UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY

Structural and functional characterization of MuB protein involved in DNA targeting for transposition

Marija Dramićanin

Madrid, June 2014

UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY

Structural and functional characterization of MuB protein involved in DNA targeting for transposition

Doctoral thesis submitted to the Universidad Autónoma de Madrid for the degree of PhD by M. Sci. in Biology,

Marija Dramićanin

Thesis Director

Dr. Santiago Ramón-Maiques

Structural Bases of Genome Integrity group

Structural Biology and Biocomputing Programme

Spanish National Cancer Research Centre

Dr. Santiago Ramón-Maiques, head of the Structural Bases of Genome Integrity Group in the Spanish National Cancer Research Centre (CNIO),

CERTIFIES

That Ms Marija Dramićanin, Master in Biology by the University of Belgrade, has completed her Doctoral Thesis **"Structural and functional characterization of MuB protein involved in DNA targeting for transposition"** and meets the necessary requirements to obtain the PhD Degree in Molecular Biosciences. To this purpose, she will defend her Doctoral Thesis at the Universidad Autónoma de Madrid. The Thesis has been carried out under my direction and hereby I authorize it to be defended to the appropriate Thesis Tribunal.

I hereby issue this certificate in Madrid on June 10th 2014.

Santiago Ramón-Maiques

PhD Thesis Director

Miguel A. Íñiguez

PhD Thesis Tutor

This thesis, submitted for the degree of Doctor of Philosophy at the Universidad Autonoma de Madrid, has been completed in the Structural Bases of Genome Integrity Group at the Spanish National Cancer Research Centre (CNIO), under the supervision of Dr. Santiago Ramón-Maiques.

This work was supported by the following grants and fellowships:

La Caixa/CNIO International PhD Fellowship. 2010 call – Marija Dramićanin

Proyecto I+D, Ministerio de Ciencia e Innovación, Spain – Dr. Santiago Ramón-Maiques

Acknowledgements

After writing all these pages, comes the part that I find the most important. It was a long and hard journey, but at the end very exciting. Journey that I was sure I want to take, but without really having an idea what I am getting myself into. I have to thank to so many people that have been with me, help and support me during these four years.

First, I would like to thank Santiago for taking a chance with me. It was not easy over the time, but we have passed trough it and I have learned very much. I think we both have. Thank you for giving me the chance to be part of a Mu project. That was a great experience for me. I have learned a lot and had an opportunity to collaborate with amazing scientists. Most importantly, working on this project, I started to really like and understand what the science is. Thank you for teaching me how to ask the questions, think, work and finally communicate my work. Although I was not lucky or maybe persistent enough to get crystals, you still decided to take your time and teach me how to solve the crystal structure. Thank you for helping me with all the presentations, posters, talks and all other things. Finally, thanks for all the encouragement, support and understanding.

Ara and Nada. Nada and Ara. I cannot talk about the two of you separately or decide who comes first. Thank you for being along me, supporting and teaching me all these years. Without the two of you, I don't think all this would be possible. Nada maybe means nothing in Spanish, but in my language it means all, it means hope. I can only hope that one day I can become a person or a scientist as you. Ara, you are one of the best people I know. I was lucky to meet you and to work with you. Alba, loca, you brought the happiness to our lab. It was amazing and inspiring to see you smiley every day! I am happy that I was able to do my PhD with you, you made me feel better every day. It is hard (even to me) to be all sad and serious when a person sees you. Thanks for all the fun we had together, for letting me be part of your wedding, meet your family, take me to Écija on carneval, all the trips we did together. Marta, we met in the lab, but became real friends outside of the science world. You gave me a Spanish family and I am very grateful for that. Thank you for all the Christmas dinners, Reyes, roscónes, summer days and nights, parties, trips, and all happy and also sad moments we have spent together. Leyre, pequeña, it was great to have in

the lab! Paco y Maria, we have just met but the little time we spent together was great!

I thank to Daniel and all the people from his group: Guille, Johanne, Luis, Chevi, Marta and Debi. Debi, thank you for being with me all these years. Thanks for all the advices, support and help you gave me. You made it all worth it. Thanks for all the trips, parties, fun we had together, thank you for taking me to Italy and introducing me to your parents and friends, letting me be part of one of the most important day for you. I was so lucky to meet you.

I also thank to Guillermo and all the people from his group: Nehar, Inés, Rafa, Lizi, Pilar, Stefano, Igor, Jesus, Pablo, Dario, Jaime, Juan, Anita and Jaska. Jaska, thank you for teaching me EM, for answering so many of my questions, supporting me with experiments and throughout my entire thesis. Jaska, ti si genijalna, sta da ti kazem! Thank you Blanca and Ramón. Blanca, maybe it wasn't such a bad luck I didn't get crystals, as I got the chance to learn about the NMR.

Milena, Sandra, Svetlana. Stavljam vas po azbucnom redu, jer kako drugacije da odlucim koja dolazi prva?! Zajedno smo prosle kroz studiranje i jos mnooogo toga. Iako ste fizicki bile daleko, sve vreme sam osecala da ste uz mene i hvala vam na tome. Jako mi nedostajete! Mici i Urose isto vazi i za vas. Daleko, a blizu. Nedostajete mi, rastavici. Iva. Hvala ti za sve. Uzasno mi je tesko da nadjem reci za ono sto zelim da ti kazem. Mislim da ti znas i sama. Dosle smo do te tacke da se razumemo i bez reci. Ali cu ipak reci jednostavno, hvala ti za sve veceri koje smo provele zajedno, za sve cemu si me naucila, svu podrsku koju si mi pruzila. Prosle smo i kroz dobe i kroz lose stvari zajedno i usput si postala deo moje porodice. Marko. Svega ovoga ne bi bilo da je...Hvala ti, za svu podrsku, savete i ljubav koju si mi pruzio. Uz tebe sam se osecala sigurno i samo razmisljanje o tebi, nama, dalo mi je snagu da prodjem kroz sve ovo. Ti si moja ljubav i uvek ces biti. Volim te. Hvala mojoj porodici, mom ujku Zoku, posebno.

Mama, tata i Ana. Volim vas. Hvala vam na svemu sto ste me naucili, kako ste me vapitali i podrzali. Hvala vam sto ste me naucili da cenim znanje, da uvek tezim ka

visem i da nikad ne odustajem. Moj svaki uspeh je vas uspeh. Ja bez vas ne bih bila to sto jesam, zbog toga sav trud i rad za ove cetiri godine posvecujem vama

Abstract

DNA transposons are ubiquitous in the genomes of all forms of life and play important evolutionary roles in generating gene diversity and in shaping genomic landscapes. Mu phage is one of the most complex and efficient DNA transposon. Mu transposition requires two phage-encoded proteins: the transposase MuA and the accessory protein MuB. MuB is an ATP-dependent non-specific DNA binding protein that regulates the activity of the MuA transposase and captures target DNA for transposition. Mechanistic understanding of MuB function has previously been hindered by its poor solubility and tendency to aggregate. We combined bioinformatic, mutagenic, biochemical, electron microscopic and NMR analyses to unmask the structure and function of MuB. We demonstrate that MuB is an AAA+ ATPase composed of an N-terminal appendage and an AAA+ ATPase module that upon ATP binding forms helical filaments on the DNA. We also identify critical residues for its ATPase, DNA binding, protein polymerization and MuA interaction activities. Using single-particle electron microscopy, we show that MuB assembles into helical filaments that coat the DNA without deforming it, resulting in a unique protein-DNA symmetry mismatch. These findings, together with the influence of MuB-filament size on strand-transfer efficiency, lead to a model in which MuB-imposed symmetry transiently deforms the DNA at the boundary of the MuB filament and results in a bent DNA favored by MuA for transposition. We have also observed the tendency of the MuB filaments to form bundles in an N-terminal appendage dependent manner. The structure of the N-terminal appendage was solved by NMR spectroscopy. The N-terminal appendage is strikingly similar to the λ-repressor like DNA-binding domains, strongly suggesting that this MuB domain could be involved in DNA recognition. This led us to propose a new model of Mu phage target immunity in which filament-filament interactions mediated by the N-terminal appendage could aid in the condensation of the phage DNA, occluding it from the transposase.

Presentación

Los transposones de ADN son ubicuos en los genomas de todos los seres vivos y juegan un papel evolutivo importante en la generación de diversidad génica y en la definición de los genomas. El bacteriófago Mu es uno de los transposones de ADN más complejos y efectivos. La transposición de Mu requiere dos proteínas codificadas por el fago: la transposasa MuA y la proteína accesoria MuB. MuB es una proteína de unión a ADN dependiente de ATP que regula la actividad de la transposasa y captura el ADN diana para la transposición. La comprensión de la función de MuB a nivel mecanístico se ha visto dificultada por su baja solubilidad y su tendencia a agregar. Hemos combinado análisis bioinformatico, mutagénesis, bioquímica, microscopía electrónica y NMR para desenmascarar la estructura y la función de MuB. Demostramos que MuB es una ATPasa AAA+ compuesta de un apéndice N-terminal y un módulo AAA+ que al unir ATP forma filamentos helicoidales sobre el ADN. También hemos identificado residuos clave para la unión e hidrólisis del ATP, unión al ADN, polimerización e interacción con MuA. Usando miscroscopía electrónica de partículas individuales mostramos que MuB se ensambla en filamentos helicoidales que cubren el ADN sin deformarlo, resultando en un mal emparejamiento único. Estos resultados, junto a resultados de cómo el tamaño de los filamentos de MuB afecta a la eficiencia de la transposición, sugieren un modelo según el cual la simetría impuesta por el filamento de MuB deforma transitoriamente el ADN al final del filamento, presentándolo como un mejor sustrato para la transposasa MuA. Hemos observado también la tendencia de los filamentos de MuB a formar haces de un modo que depende del apéndice N-terminal. Hemos resuelto la estructura del apéndice N-terminal mediante espectroscopía de RMN. El apéndice N-terminal es sorprendentemente similar a los dominios de unión de ADN de la familia del represor λ, sugiriendo que este dominio de MuB podría estar implicado en el reconocimiento del ADN. Estos resultados nos llevan a proponer un nuevo mecanismo de inmunidad del fago Mu, en el que las interacciones entre filamentos mediadas por el apéndice N-terminal podrían ayudar a la condensación del genoma de Mu, ocultándolo así de la acción de la transposasa.

Table of contents

Abbreviations

AAA+	ATPases Associated with diverse cellular Activities
ATP	Adenosine triphosphate
ATPγS	Adenosine 5'-(γ-thio)-triphosphate
BAF	Barrier to Autointegration Factor
CAP	Catabolite Activator Protein
CD	Circular dichroism
DNA	Deoxyribonucleic acid
cDNA	Complementary DNA
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
EGFP	Enhanced Green Fluorescent Protein
FF	Fast flow
GFP	Green fluorescent protein
GST	Glutathione S-transferases
HIV	Human immunodeficiency virus
HSQC	Heteronuclear Single Quantum Coherence
HTH	Helix-turn-helix
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Broth growth media
LTRs	Long Terminal Repeats
Mu	Mutator
MALS	Multi Angle Light Scattering
MLV	Moloney Leukemia Virus
MW	Molecular weight
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser effect spectroscopy
NOE	Nuclear Overhauser effect
NTA	N-terminal appendage
OD	Optical density
PAGE	Polyacrylamide Gel Electrophoresis

PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Base
PEI	Polyethyleneimine
PMSF	Phenylmethylsulfonyl fluoride
RNA	Ribonucleic acid
RPM	Revolutions per minute
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase
SDS	Sodium dodecyl sulfate
TIRF	Total internal reflection fluorescence microscope
TLC	Thin Layer Chromatography
Tn	Transposon
TOCSY	Total Correlation Spectroscopy
UTR	Untranslated region
WT	Wild type

Introduction

1. Transposons, the jumping gene machines

Transposition is a specific form of DNA recombination used by certain DNA elements to move from one site of the genome to another. These mobile genetic elements are called transposable elements, transposons or simply "jumping genes". Barbara McClintock discovered transposable elements more than fifty years ago (McClintock, 1950). Today, we know that they are ubiquitously present in all the species studied and DNA sequence analyses have shown that they can be highly abundant, making more than 40% of the human (Lander et al., 2001), mouse (Waterson et al., 2002) or rice (Goff et al., 2002) genomes.

Transposons play an important role in structure and evolution of genes and genomes from bacteria to humans (Berg and Howe, 1989; Craig et al. 2002). They drive the evolutionary changes by causing inversions and deletions of chromosomal DNA, which can be either a direct result of intramolecular transposition, or by providing dispersed regions of homology that can be recognized by the DNA recombination machinery of the host. Transposition of some elements can lead to the transduction of flanking DNA that causes the rearrangements of host genes (Moran et al., 1999; Mendiola et al., 1994). They also distribute number of determinants such as antibiotic resistance or virulence factors (Scott, 2002).

2. There are three classes of transposons

Based on their overall organization and a mechanism of transposition, transposons can be divided into three families: DNA transposons, viral-like retrotransposons and poly-A or non-viral retrotransposons (Figure 1). DNA transposons are in the form of DNA throughout the whole recombination process and both types of retrotransposons move to a new DNA location using an RNA intermediate.

DNA transposons are delimited on both terminal flanks by inverted repeat sequences, which are the recombinase recognition sequences, and contain at least one gene encoding for a transposase (Figure 1). The transposase (or integrase) is the enzyme responsible to cleave the ends of the transposon, freeing it from its initial

location in the genome, and also for cleaving the DNA target site where the element is inserted. DNA transposons may carry additional genes, some of which encode accessory proteins that are often necessary for successful transposition.

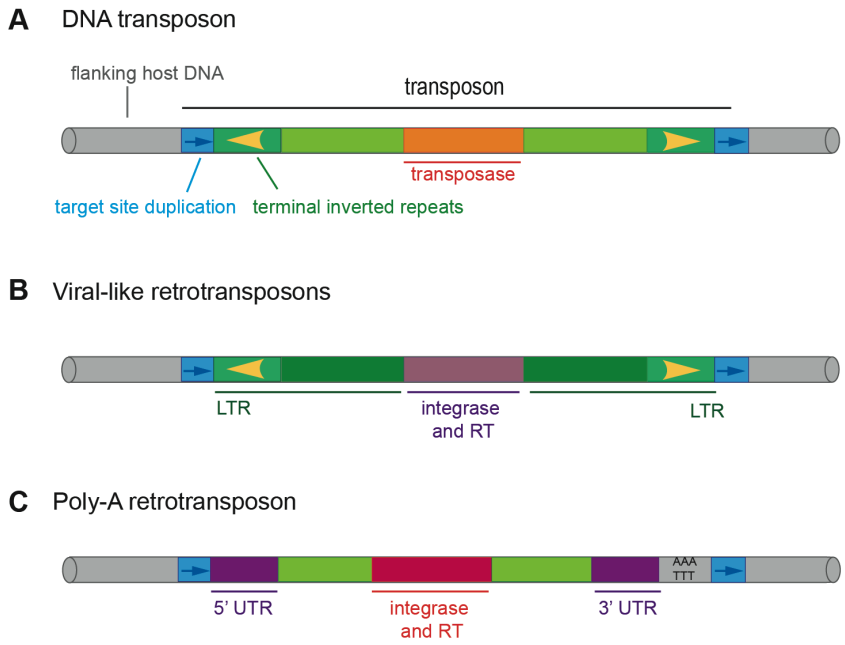


Figure 1. Three classes of transposons and their genetic organization. (A) DNA transposons. The element includes inverted repeat sequences (yellow arrows), which are the recombination sites, and a gene encoding transposase. (B) Viral-like retrotransposons. The element includes two long terminal repeats (LTR) and region encoding integrase and reverse transcriptase (RT). (C) Poly-A retrotransposons. The element ends with 5' and 3' UTR sequences and encodes for integrase and reverse transcriptase.

Viral-like retrotransposons have a terminal inverted repeats that are embedded within longer sequences called long terminal repeats or LTRs. These transposons also encode for the two enzymes needed for their mobility: integrase (transposase) and reverse transcriptase (RT). RT is an enzyme that converts RNA template to a complementary DNA (cDNA) in a process called reverse transposition.

Unlike the other two groups, Poly-A or non-viral retrotransposons do not have terminal inverted repeats. Instead, these transposons have two ends with different sequences: 5' UTR (untranslated region) and 3' UTR, followed by poly-A sequence, which is a stretch of AT base pairs. They also have genes encoding for integrase and RT (Watson et al, 2004).

3. Transposition mechanism

Mobile genetic elements share a similar general mechanism to insert their DNA into a new DNA target site. The understanding of this mechanism comes mostly from the studies on bacteriophage Mu, the first DNA mobile element for which a defined *in vitro* transposition system was available (Mizuuchi, 1983). General model of transposition involves the initial binding of the transposase to the terminal inverted repeats flanking the mobile element. Then, the transposase brings the two ends in close proximity forming a stable protein-DNA complex known as synaptic complex or transposome (Suretter et al., 1991; Mizuuchi et al., 1995,). The transposase catalyzes the cleavage of one strand at each of the two junctions between the transposon sequence and the host DNA. This release two 3'-OH groups at the ends of the transposon sequence that attack the DNA phosphiester bonds at the target DNA site, in a reaction called DNA strand-transfer. The transposome ensures that the two strands attacked by the 3'-OH groups are separated by a short spacing of a few nucleotides. This distance is fixed and characteristic of each transposon and generates a typical short duplicated sequence that flank the new copies of the element. Following the strand-transfer reaction, the transposome disassembles and the cellular machinery completes the remaining steps.

Some DNA transposons move by a mechanism in which the 5' ends at the junction with the host DNA are also cleaved. This step can be carried out by the transposase alone or with the help of an accessory protein (e.g. TnsA and TnsB in transposon Tn7) (May and Craig, 1996; Sarnovsky et al., 1996). This mechanism involves the complete excision of the transposon and the insertion into a new DNA site, and it is called, therefore, cut-and-paste transposition.

Mu phage uses a mechanism called replicative transposition catalyzed by two phage-encoded proteins: transposase MuA and accessory protein MuB (Figure 2). In Mu phage replicative transposition, only one strand at each end of the transposon is cut and joined to the target DNA site (Craigie and Mizuuchi, 1985; Lavoie and Chaconas, 1996). This generates a DNA intermediate where 3’ ends of the phage stay attached to the target DNA while 5’ ends remain joined with the flanking DNA, resulting in a

formation of a double-branched DNA structure. The host cell replication machinery assembles at the end of these intermediate generating two copies of the phage. One end of each copy of MuB genome stays attached to the donor site and the other end to the target site. This structure, named cointegrate, is resolved by subsequent recombination between the element copies to generate two plasmids, one a target plasmid containing a simple insertion and the other a copy of the original donor plasmid (Berg and Howe, 1989).

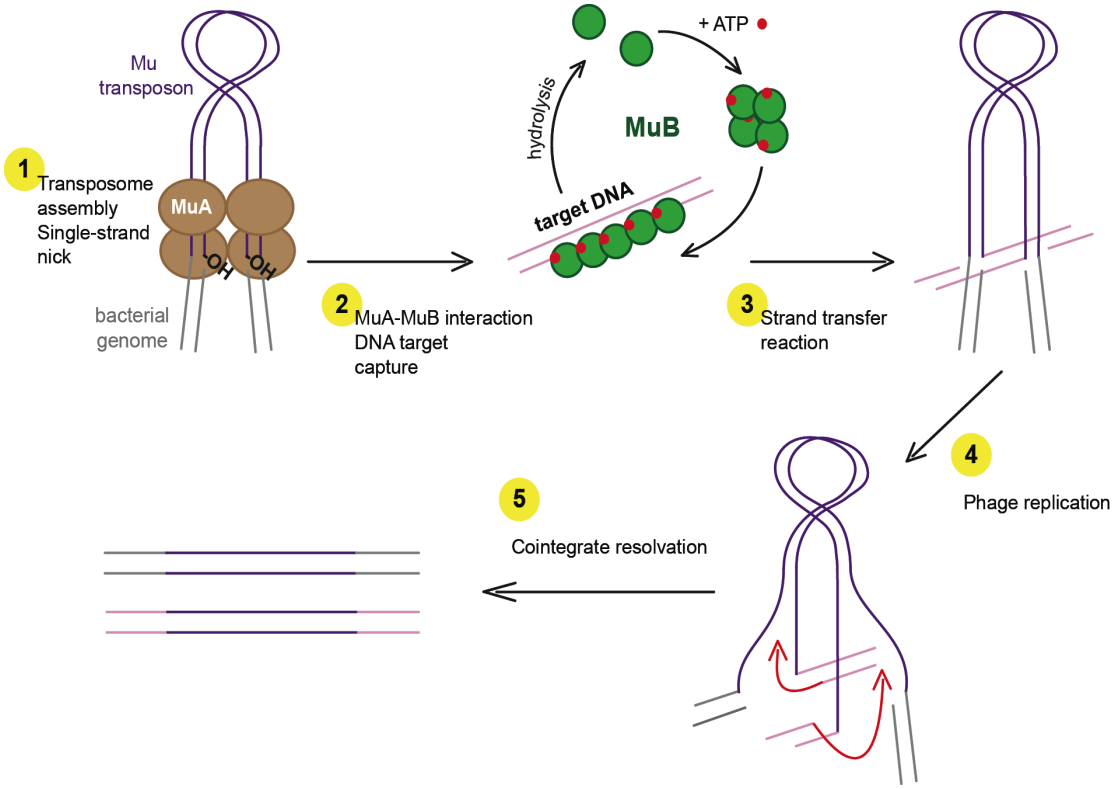


Figure 2. Mu phage replicative transposition.

Retroviruses and viral-like retrotransposons use the same steps of DNA cleavage and strand-transfer as those described for cut-and-paste DNA transposons. However, the transposition cycle starts with transcription of the retrotransposon into RNA, followed by reverse-transciption by the enzyme reverse transcriptase to generate a double stranded cDNA freed from the flanking host regions, which will be inserted by an integrase into the DNA target site (Boeke and Stoye, 1997).

The cleavage of the 3' ends and the strand transfer reaction are common for common for transposases and for retroviral integrases. The crystal structures of the

catalytic domains of HIV integrase (Dyda et al., 1994), RSV integrase (Chook et al., 1994), Tn5 transposase (Davies et al., 2008), and MuA transposase (Montaño et al., 2012) show that, despite the lack of sequence similarity, their catalytic domains are largely superimposable and that all three enzymes belong to the DDE-motif transposase/integrase protein family (Hickman et al 2010). The defining characteristic of these proteins is the invariant DDE motif, which was shown to be essential for transposition both *in vivo* and *in vitro* (Haren et al., 1999; Craig et al., 2002). Revealing the common features in transposition reaction and between transposition and other types of recombination such as the breakage and joining reactions in V(D)J recombination that mediate immunoglobulin gene assembly (Craig, 1996; van Gent et al., 1996) has led to understanding of the molecular links between many systems whose strong biological connections have long been appreciated (Bukhari, 1977). Thus, development of *in vitro* Mu phage transposition reaction have provided precious tool for the understanding of a wide variety of biological processes in many different organisms. This includes development of HIV integration assays (Craigie et al., 1991) and studies of development of the immune system in vertebrates.

4. Phage Mu is a gigantic transposon

The name of the phage Mu comes from the ‘’mutator‘’, as impressively high rate of Mu transposition causes accumulation of new mutations into the host cells due to insertion of the phage into cellular genes. Mu genome is about 37 kb and includes genes that encode proteins for the regulation of transposition, lysis of the host cells and phage head and tale formation (Morgan et al., 2002). The life cycle of Mu phage is shown in Figure 3 A. When Mu infects a sensitive host, the linear DNA enters the cell and the Mu DNA is inserted into the recipient genome via a non-replicative, "cut and paste" mechanism. After the cell enters into lytic cycle, the transposition proteins get expressed and Mu transposes by a replicative mechanism into new sites on the chromosome (Watson et al., 2004).

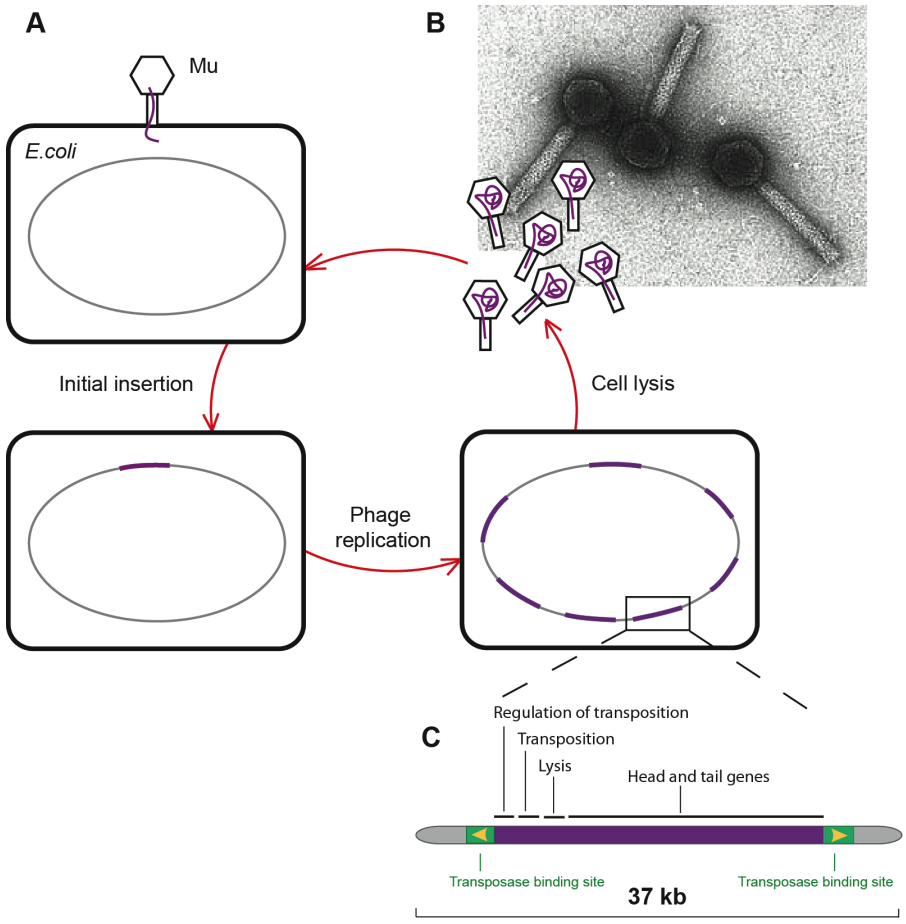


Figure 3. Mu phage. (A) Mu phage life cycle. (B) Electron micrographs of purified Mu phage particles (taken from Grundy and Howe, 1984). (C) Mu phage genome organization.

Mu completes up to 100 rounds of replicative transposition during a single growth cycle, which means that by the end of this process, the Mu DNA content reaches nearly the size of the host. This makes Mu phage one of the most complex and efficient transposable elements known (Mizuuchi, 1992). Meanwhile, late phage gene products are made (including phage heads, tails, lysis proteins, etc). After the assembly of the new phage particles, the host is lysed, releasing 50-100 new phages (Watson et al., 2004).

As it was mentioned before, efficient Mu phage replicative transposition involves two Mu-encoded proteins: MuA transposase and accessory protein MuB. MuA transposase is a very well characterized, both structurally and functionally. MuA is a 75 kDa protein that can be divided into four domains: N-terminal domain, transposon binding domain, catalytic core and C-terminal domain. The N-terminal 76

residues long domain is required for efficient assembly of MuA transposome. Following domain, that involves residues 77-247, is responsible for the recognition of the Mu phage terminal inverted sequences (Nakayama et al., 1987; Leung et al., 1989). The core domain (residues 248-574) forms the catalytic center of the transposase and also has some non-specific DNA-binding activity (Baker and Luo, 1994; Nakayama et al., 1987). Finally the C-terminal domain (residues 575-663) of MuA has a region important for the protein tetramerization and interaction with the MuB protein (Wu and Chaconas, 1994). Mu phage transposition requires the tetramerization of the MuA protein in a transposome, whose structure was recently solved by the group of Phoebe Rice (Motaño et al., 2012). Interestingly, the structure shows the target DNA that is highly bent. It has been previously described that some other transposones also show strong preference for the bent target DNA (Milot et al., 1994; Pribil et al., 2003). However, Mu transposition is inefficient in the absence of the second Mu phage transposition protein, MuB. The reason is because MuB controls the activity of the transposase at several levels. First, MuB assists in the assembly of the MuA tetramer at the transposon ends (Mizuuchi et al., 1995). The binding of MuB to MuA also prevents premature disassembly of the transposome by blocking the recognition of the C-terminus of MuA by the chaperone ClpX (Levchenko et al., 1997). Additionally, MuB activates MuA allosterically to efficiently carry out the strand transfer reaction (Baker et al., 1991). Finally, MuB is needed for the selection of the proper target DNA for the transposition, since in the absence of MuB, MuA utilizes only intra-molecular DNA sites for transposition, leading to self-destruction (Craigie et al., 1985; Maxwell et al., 1987; Adzuma and Mizuuchi, 1988).

5. DNA target selection and transposition immunity

Typically transposons exhibit little sequence selectivity in their choice for the target DNA. As a result, transposons can jump within genes and disrupt their function. However, some transposons show a target site selection that facilitates element spread and optimizes the element-host relationship (Craig et al., 2002). The Tn7 transposon, for example, does not insert into essential genes, but rather into a

specific site in bacterial chromosomes where insertion will not harm the host bacterium. Transposons Ty1 and Ty3 insert upstream of promotors, while others like R2 avoid inactivation of essential genes by targeting ones that have many copies. However, element insertion into particular safe sites is certainly not wide spread because many successful elements, like phage Mu, insert at many different target positions, occasionally moving into essential genes (Craig, 1997).

As mentioned before, Mu phage efficiently amplifies its genome reaching nearly the size of the host genome by the end of the single growth cycle. Thus, the problem for Mu element is not how to avoid inserting into essential bacterial genes, but how to choose the target DNA site in order to avoid integration into itself, a situation that would lead the element to self-destruction. Mu, as several other transposons, like Tn3 and Tn7, efficiently avoid self-destructive insertion by a phenomenon called "target immunity" (Robinson et al., 1977; Lee et al., 1983; Hauer and Shapiro, 1984).

In vitro studies with phage Mu provided the first molecular insights into the targert immunity phenomenon (Adzuma and Mizuuchi, 1988). Interplay between MuA transposase and the MuB protein is in the center of this mechanism. MuA-MuB stimulated transposition displays target immunity, avoiding target sites near pre-existing Mu-end sequences where MuA binds. This is because MuA stimulates the ATPase activity of MuB and since this interaction takes place more frequently when the two proteins are bound on the same DNA molecule relatively close to each other, MuA triggers the dissociation of MuB from DNA sites near the Mu ends more efficiently than those at a distance (Han and Mizuuchi, 2010). Thus, by the time MuA assembles the active complex with synapsed Mu DNA ends ready to capture target DNA, MuB has been depleted from their vicinity and accumulates on DNA sites away from the Mu end sequence becoming a preferable site for the transposition. This explains the strong preference of the transposing Mu to target DNA sites at substantial distances away (5-25 kb) from its original site (Manna and Higgins, 1999).

The group of Rasika Harshey proposed the new mechanism that protects Mu from self-integration named ‘’Mu genome immunity‘’ (Ge et al., 2010). Contrary to what is expected in target immunity mechanism, Harshey’s group reports strong MuB binding throughout the Mu genome. While the target immunity is functional outside of Mu phage ends, level of protection offered by this mechanism is insufficient to

explain the protection seen inside Mu. Alternative immunity mechanism proposes that Mu genome is segregated into an independent ''Mu domain'' that may be assisted by specific Mu sequences and other nucleoid-associated proteins, stimulating polymerization of MuB on the Mu genome to form a barrier against self-integration (Saha et al., 2013). Similar form of immunity against self-integration is present in some retroviruses such as Moloney leukemia virus (MLV) and Human immunodeficiency virus (HIV). These viruses protect their DNA against intramolecular insertion by using a protein called barrier to autointegration factor (BAF). BAF is a non-specific, DNA-binding protein that bridges viral DNA and condenses it (Zeng et al., 2000; Bradley et al., 2005). BAF protein plays a dual role. On one hand, it compact the DNA reversibly to prevent auto-integration, while on the other it promotes intermolecular target capture (Suzuki et al., 2002).

6. Knowns and unknowns of MuB protein

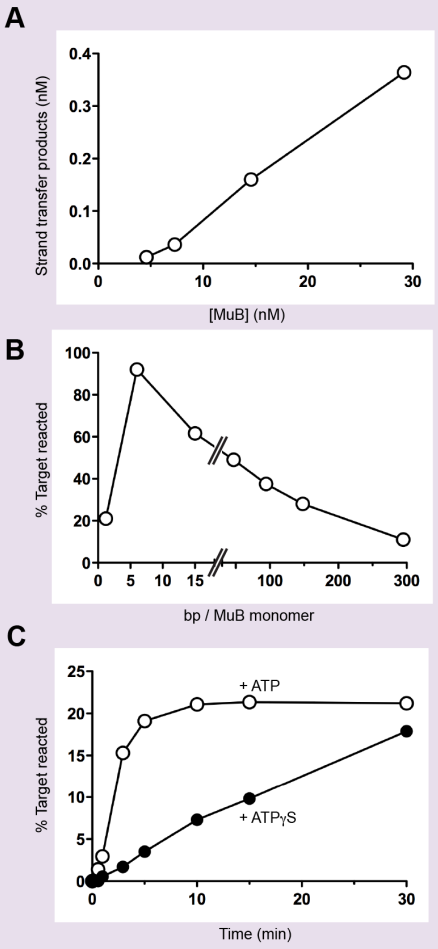
Although structural analysis of MuA protein has advanced (Clubb et al., 1994; Rice et al., 1995; Schumaker et al., 1997), the available structural information on MuB is limited. MuB is a small (35 kDa) ATP dependent non-specific DNA binding protein with relatively low ATPase activity. (Miller et al., 1984; Chaconas et al., 1985; Maxwell et al., 1987). The domain organization of MuB was first studied by limited proteolysis. These studies shown that the 312 amino acid MuB polypeptide can be divided into two pieces: a large 25 kDa N-terminal and a smaller 10 kDa C-terminal fragment (Teplow et al., 1988). The N-terminal fragment was shown to bind and hydrolyze ATP and contains Walker A and Walker B motifs (Yamauchi and Baker, 1998). It was shown that this part of the protein can bind ATP even in the absence of the C-terminal fragment (Chaconas, 1987, Teplow et al., 1988). The N-terminal fragment also shows the ability to bind the DNA (Milliner and Chaconas, 1998). C-terminal fragment of MuB is exceedingly insoluble with a high propensity to aggregate. However, the group of George Chaconas manage to determine the structure of this fragment by NMR spectroscopy in a high salt condition (1.5M NaCl) (Hung et al., 2000). The structure revealed that that the C-terminal fragment of MuB

is a four helical bundle. It also identified potential DNA binding properties and revealed a resemblance with the N-terminal domain of *E. coli* replicative helicase DnaB. Beyond this, no sequence similarity between MuB and other proteins has been reported. Its tendency to aggregate in the presence of ATP hampered efforts at structural characterization (Chaconas et al., 1985; Teplow et al., 1988).

It is known that upon ATP binding MuB polymerizes preferentially on DNA, but in the absence of DNA it still can form polymers of variable sizes (Adzuma and Mizuuchi, 1991; Greene and Mizuuchi, 2002a). Kinetic studies showed that MuB oligomerization stimulates the ATPase activity (Adzuma and Mizuuchi, 1991). In the presence of DNA, MuB-ATP oligomers bind DNA without sequence specificity yet with a preference for A/T-rich regions (Miller et al., 1984; Greene and Mizuuchi, 2002a; Greene and Mizuuchi, 2002b). MuB polymers were first observed by Total Internal Reflection Fluorescence Microscopy (TIRF). For this purposes, MuB was labeled with an Enhanced Green Fluorescent Protein (EGFP) at the N-terminus. These experiments showed that, in the presence of the ATP, MuB binds along the DNA molecule forming many short separate segments of polymers. The more of the EGFP-MuB is added the protein-covered segments elongate to form an apparently continuous polymer that fully coats the DNA. On the other hand, the hydrolysis of ATP reverses this process, triggering disassembly of the MuB polymer (Greene and Mizuuchi, 2004; Greene and Mizuuchi, 2002b; Greene and Mizuuchi, 2002c).

To measure to what extent the presence of MuB on the DNA could affect transposition, the group of Kiyoshi Mizuuchi tested the effects of the increase of the MuB concentration on an efficiency of the MuB-stimulated strand transfer reaction (Box 1). Mizuuchi's group shows that the increase in the concentration of MuB positively affects the selection of the DNA as a target for the transposition. However, DNA that is fully saturated by the MuB is shown to be refractory to the transposition. Prior to the beginning of this thesis, MuB aggregates were visualized by negative staining electron microscopy (EM) and it revealed that MuB forms filaments in the presence of the ATP.

Box 1



The extent of MuB binding controls the selection of target DNA for MuA transposase

In order to assess how much of MuB is needed to activate the strand transfer step of transposition, the group of **Kiyoshi Mizuuchi** asked how the MuB concentration or stoichiometry with DNA affects the efficiency of transposition. At MuB concentrations below 10 nM, the target DNA strand transfer efficiency was sigmoidal in response to the increase of MuB concentration, in agreement with the necessity of MuB oligomerization for DNA binding (A). At moderately higher MuB concentrations, the reaction efficiency increased roughly linearly with MuB concentration, producing one transposition product per ~50-100 MuB monomer. The higher the density of MuB on DNA, the more efficiently the MuB-bound DNA fragment was used as target. However, when DNA was completely saturated by bound MuB, it suddenly became a poor strand transfer target and only 20% of the DNA was used for transposition (B). The inhibition of transposition by saturating MuB concentration was more severe when ATP was substituted by ATPγS (C).

The lack of structural information about the MuB protein made it difficult to understand how ATP controls MuB polymerization, what kind of polymer MuB forms, how MuB polymer interacts with DNA, nor how MuA interacts with MuB polymer and triggers ATP hydrolysis and DNA dissociation. In this thesis, I set to answer these questions in order to better understand the structure and the function of MuB protein. This work was done in collaboration with groups of Kiyoshi Mizuuchi and Alasdair Steven from the National Institute of Health, Bethesda, United States. Using MuB wild type and mutant proteins produced as a part of this thesis, Kiyoshi Mizuuchi's group has preformed Mu phage *in vitro* transposition assays and Naoko Mizuno did a 3D cryo-EM recostructions and molecular docking experiments. As all these results were integral part of this thesis, their summaries are presented as purple boxes.

Objectives

Objectives

The general aim of this thesis is to describe the structural features of the MuB protein to unravel their connections with the functional aspects of this protein in Mu phage DNA target selection. This global aim implies attaining the particular objectives:

- Characterization of the ATP-dependent MuB filament formation on the DNA.
- Understanding the mechanism by which MuB selects the DNA for transposition and confers target immunity.

Objectivos

Objectivos

El objetivo general de esta tesis es describir las características estructurales de la proteína MuB para desentramar sus conexiones con los aspectos funcionales de esta proteína en la selección del ADN diana por el fago Mu:

- Caracterización de la formación de filamentos de MuB dependientes de ATP en el ADN.
- Entender el mecanismo por el cual MuB selecciona el ADN para la transposición y confiere inmunidad.

Materials and methods

1. Cloning

The plasmid pET14b-MuBwt (Yamauchi, 1998) was used to express a His_6 -tagged MuB fusion protein with a thrombin cleavage site between the His-tag and the MuB protein sequence. For removal of His-tag by PreScission, instead of thrombin, MuB coding sequence was amplified by PCR using the forward primer 5'-AAGTTCTGTTTCAGGGCCCGATGAATATTTCCGATATTCG-3' and reverse primer 5'-ATGGTCTAGAAAGCTTTAATTACGCAGCAGCGTTG-3' and cloned into pOPIN-B (Oxford Protein Production Facility UK) linearized with KpnI and HindIII, using the In-Fusion cloning technology (Clontech) (Berrow, 2007). This fast and directional cloning technique is based on the 15 bp overlap (underscored primer sequence) between the amplified DNA fragment and the digested vector. Although not a prerequisite, the amplified MuB gene and the digested vector were gel-purified prior to the In-fusion reaction with the QIAquick Gel Extraction Kit (Qiagen), to improve the efficiency of the cloning. The resulting plasmid was verified by sequencing and transformed into E. coli BL21 (DE3) pLysS cells (Novagen). this construct was named pOPIN-B-MuBwt.

2. MuB protein expression and purification

BL21 (DE3) pLysS E. coli cells (Novagen) transformed with the pET14b-MuBwt or with the pOPIN-B-MuBwt plasmid were grown at 37 °C in 1 L of LB medium supplemented with 100 µg ml^{-1} ampicillin or 50 µg ml^{-1} kanamycin, respectively, and 34 µg ml^{-1} chloramphenicol. When the cultures reached mid-exponential phase (OD_{600} =0.6-0.8), protein expression was induced by addition of 0.8 mM isopropyl-D-thiogalactopyranoside (IPTG) and incubation was continued at 37°C. After 4 hours, the cells were harvested by centrifugation, resuspended in buffer A (20 mM Tris-HCl pH 8.0, 1 M NaCl, 10 % glycerol, 2 mM β-mercaptoethanol) supplemented with 0.2 mM phenylmethanesulphonyl fluoride (PMSF) and Complete EDTA-free protease inhibitor cocktail tablets (from Roche) and disrupted by sonication on ice using a cycle of 3 s on, 1 s off at 37 % amplitude for a total of 10 min with breaks after every

2.5 min. The lysate was clarified by centrifugation in a Beckman Ti-45 rotor at 40 000 rpm min^{-1} for 40 min. The supernatant was filtered trough a 0.45 µm pore filter (JET BIOFIL) and applied onto a 5 ml Ni^{2+}-loaded HisTrap Chelating FF column (GE Healthcare) in buffer A using Peristaltic pump P1 (GE Healthcare). Following extensive washing of the column with 100 ml of buffer A containing 30 mM imidazole, the protein, which was tagged at the N-terminus with His6, was eluted by increasing imidazole concentration to 300 mM stepwise with the AKTA FPLC system. To prevent protein degradation, EDTA was added to the fraction collection tubes prior to protein elution so that the final concentration of EDTA in the eluted sample was 5 mM. The sample was diluted in buffer B (30 mM Hepes-Na pH 7.6, 0.15 M NaCl, 20% glycerol, 5 mM EDTA and 1 mM DTT) to reduce the ionic strength. A significant fraction (~40%) of the protein precipitated and was removed by centrifugation and filtration. Thrombin (from human; Sigma) was added to the soluble fraction at a final concentration of 0.05 U μL^{-1}, and cleavage of His-tag was complete after 1.5 hours incubation at 4 °C. The sample was applied in two batches to a 5 ml HiTrap SP FF column (GE Healthcare) equilibrated in buffer B, and MuB was eluted by a salt gradient at ~250 mM NaCl with AKTA FPLC system. After increasing NaCl concentration to 1M, the eluted sample was concentrated to 8.6 mg/ml by centrifugal ultrafiltration using a Centricon YM-10 of 10 kDa cutoff (Millipore), flashed frozen in liquid nitrogen and stored at -80 °C. A Superdex 75 gel filtration column, equilibrated in 20 mM Tris-HCl pH 8, 1 M NaCl, 10% glycerol, 1 mM EDTA, 1 mM dithiothreitol (DTT), was included as a last purification step.

For the MuB constructs with PreScission cleavage site, the above procedure was followed except the thrombin digestion step was replaced with PreScission protease digestion (1:20, protease to MuB mass ratio) at 4 °C for 4-8 h.

3. Site-directed mutagenesis

Mutagenesis was carried out by using overlap extension in PCR reaction as previously described (Ho, 1989). Pairs of mutagenic and flanking primers (Table 1) were incorporated in independent nested PCRs to ultimately combine them in the

final product. The reaction uses flanking primers (described above in cloning section) on either end of the target sequence, plus two internal primers that contain the desired mutation, in a first round of PCR that yields two products that are gel-purified. Then, the PCR products are mixed together and because the mutagenic primers have complementary sequence, the two fragments will hybrize in the second PCR where only the flanking primers are used. The final product containing the mutated sequence was gel-purified and inserted into linearized pOPIN-B using In-Fusion (Berrow, 2007). The expression and purification of the mutants was preformed as for the MuBwt, except that His_6 tag was not removed to speed up the process.

Table 1. Oligonucleotides used in MuB mutagenesis.

Mutant	**Location in the protein**	**Primer**	**Sequence (5' to 3')**
MuB ΔC	Truncation of C-terminal helical bundle	forward	gcaaagcgcactgcaatctaataaaaccaaaaaagccg
		reverse	cggctttttggttttattagattgcagtgcgctttgc
MuB ΔN	Truncation of N-terminal appendage	forward	aagttctgtttcagggcccggcagaactacctgaacc
		reverse	atggtctagaaagctttaattacgcagcagcgttg
NTA	N-terminal appendage	forward	aagttctgtttcagggcccggcagaactacctgaacc
		reverse	atggtctagaaagctttatgcatgatatttttccagc
N202A	Sensor-I	forward	ggccttgtgctgatgggagcgcaccgggtttattcaaat
		reverse	atttgaataaacccggtgcgctcccatcagcacaaggcc
R268A	Sensor-II	forward	ccaggtgcgcttgcgattctgaatcattcac
		reverse	gtgaatgattcagaatcgcaagcgcacctgg
R187A	Intersubunit surface	forward	gttctggaagaactcgccctgttacag gaatcaacc
		reverse	ggttgattcctgtaacagggcgagttcttccagaac
R220A	Intersubunit surface	forward	acggttgaatttgccgctctgttttcccgtattgcaaagcg
		reverse	aatacgggaaaacagagcggcaaattcaaccgttctgttacc

R224A	Arginine-finger	forward	gcccgtctgttttccgctattgcaaagcgcactgcaattaataaaacc
		reverse	agtgcgctttgcaatacgggaaaacagacgggcaaattcaacc
R150A/ R151G/ K152A	Loop-1	forward	ggaatgaatgacgcaccagccggagcagggccgctctcccgc
		reverse	gcgggagagcggccctgctccggctggtgcgtcattcattcc
R151A/ K152A	Loop-1	forward	gacgcaccacgcgctgcagggccgctctcccgc
		reverse	gcgggagagcggccctgcagcgcgtggtgcgtc
R150A	Loop-1	forward	ggaatgaatgacgcaccagcccgtaaagggccgctctcccgc
		reverse	gcgggagagcggcccttacgggctggtgcgtcattcattcc
R151A	Loop-1	forward	gaatgacgcaccacgcgctaaagggccgctctcccgc
		reverse	gcgggagagcggcccttagcgcgtggtgcgtcattc
K152A	Loop-1	forward	gacgcaccacgccgtgcagggccgctctcccgc
		reverse	gcgggagagcggccctgcacggcgtggtgcgtc
K233A/ K235A/ K236A	positive cluster in linker region	forward	gcgcactgcaattaatgcaaccgcagcagccgatgttaaggctattgcggatg
		reverse	catccgcaatagccttaacatcggctgctgcggttgcattaattgcagtgcgc

4. Limited trypsine digestion

Limited proteolysis experiments were carried out by treating MuB with trypsin (from SIGMA). Enzymatic reactions were performed at room temperature in gel-filtration buffer (see above) in a final reaction volume of 15 µL. MuB at 1 mg ml^{-1} was incubated with increasing concentrations of trypsin (0.44, 2 and 10 µg ml^{-1}) for 1 h. The reaction was stopped by addition of Laemli gel loading buffer and the extent of the proteolysis was analyzed by SDS-PAGE and coomassie staining.

5. Negative staining electron microscopy

MuB was diluted to 0.07 mg/ml (2µM) in 30 mM Tris-HCl pH 8.0, 0.15 M KCl, 5mM $MgCl_2$ and 1mM DTT and incubated with 1 mM ATP or ATPγS. After an incubation time between several seconds to 5 min, 5 µL of the sample were applied to glow-discharged carbon-coated copper grids, washed with five drops of dionized water and stained with two drops of 1% uranyl acetate. To assemble MuB filaments on DNA, we used a pGEX-6p plasmid with a 1.58 kb fragment inserted between the BamHI and XhoI sites and digested with BamHI, XhoI, AlwNI and EcoRV (New England Biolabs) to generate four linear DNA fragments of approximately 0.3, 1.5, 1.7, and 1.8 kb. Alternatively, gel-purified PCR product of approximately 1 kb was used. To selectively obtain MuB filaments on the DNA the salt concentration in the buffer was increased to 0.3 M. The grids were viewed at room temperature with a Tecnai G2 Spirit electron microscope (FEI, Netherlands) equipped with a La_{B6} filament and operated at 120 kV. Micrographs were recorded on a TemCam-F416 4k x 4k pixel camera (TVIPS GmbH, Gauting, Germany).

6. ATPase activity assays

ATPase assays were carried out as described (Greene and Mizuuchi, 2002c). Reactions contained 1.7 µM MuB (or higher concentrations for the mutants), 25 mM Tris-HCl pH 8, 0.15M NaCl, 10 mM $MgCl_2$, 20% glycerol, 1mM DTT, 0.1 mg/ml bovine serum albumin and 1mM [α-^{32}P]ATP (4 mCi/µmol). At different time points, 10 µl aliquots were combined with 1.1 µl of 0.5 M EDTA to stop the reaction. Reaction products were separated on polyethyleneimine (PEI) TLC plates in 0.75 M KH_2PO_4 and quantified using a phosphoimager (Typhoon Trio - GE Healthcare) and ImageQuant software (GE Healthcare). Data was analyzed with Prism 4.0. The inhibitory effect of DNA on the ATPase activity was assayed adding 1.7 µM of a 36-mer double-stranded DNA. The effects of MuA and MuA_{trun} on the ATPase activity were measured at final protein concentrations of 0.45 µM MuB and and 6 µM MuA, in the presence of 10 µg ml^{-1} pMK589 DNA, which carries a pair of Mu end

sequences in wrong orientation (Greene et al., 2002a) and the reaction buffer contained additional 42 mM KCl.

7. Nucleotide binding assay

Centrifugal gel filtration (spin-column) assay was carried out essentially as described (Penefsky, 1977; Ramón-Maiques et al., 2002). 50 µl samples of 30 µM MuB in 20 mM Tris-HCl pH 8, 0.3 M NaCl, 10 % glycerol and 5 mM $MgCl_2$ were incubated with 1 mM α-^{32}P ATP for 0.5 min, applied to tuberculin syringes filled with 1 ml SephadexG-50 (fine) equilibrated in the same buffer, and immediately centrifuged. The unbound nucleotide is retained on the column, while the protein-bound nucleotide in the effluent was quantified using a scintillation counter. Protein recovery in the column effluent was measured by Bradford method (Bradford, 1976). The same DNA used in ATPase activity assays was included at a final concentration of 5 µM to test the effect on nucleotide binding.

8. Expression and purification of MuB N-terminal appendage

MuB N-terminal appendage (NTA) was expressed in *E. coli* BL21 (DE3) pLysS cells (Novagen) by incubation in LB medium supplemented with 50 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol at 37 °C with shaking at 220 rpm. Expression of the NTA was induced when cultures reached OD_{600}= 0.6 with 0.8 mM isopropylthio-β-galactoside (IPTG) followed by incubation with shaking at 220 rpm for 4 h at 37 °C. ^{15}N labeling was preformed by growing cells in M9 minimal medium containing 1g l^{-1} of ^{15}N ammonium chloride as the nitrogen source, following the expression by the IPTG induction as described above. Alternatively, unlabeled protein was expressed by auto-induction in ZY-5052 medium (Studier, 2005) at 37 °C with shaking at 220 rpm. When cell density reached OD_{600}= 0.6 temperature was decreased to 25 °C and cells were grown over night with shaking at 220 rpm. The cells were harvested by centrifugation at 4000 g at 4°C for 20 min, washed with PBS and stored at -80 °C. The bacterial pellet was thawed and resuspended in buffer A (20 mM Tris-HCl pH 8, 500

mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol) with 2 mM phenylmethanesulfonyl (PMSF). Lysis was performed by sonication on ice using a cycle of 3 s on, 1 s off at 37% amplitude for a total of 10 min with breaks after every 2.5 min. The lysate was then clarified by ultracentrifugation in a Beckman Ti-45 rotor at 40 000 rev min^{-1} for 40 min. The NTA was loaded on a 5 ml HisTrap column (GE Healthcare) with Peristaltic pump P1 (GE Healthcare). Then, the column was attached to an AKTA FPLC and after extensive column washing with 100 ml of buffer A plus 60 mM imidazole, MuB was eluted in buffer A plus 0.3 M imidazole. Excess imidazole was removed and the His_6 tag was cleaved by overnight or 3.5 hours dialysis against the buffer Q (20 mM Tris-Hcl pH 8, 75 mM NaCl, 1 mM DTT) with the inclusion of GST-tagged PreScission protease (0.05 mg/ml) within the dialysis bag. The dialyzed and cleaved sample was loaded onto a 5 ml Q trap FF column (GE Healthcare, USA) with AKTA FPLC system to retain the protein. Protein eluted by increasing NaCl concentration to 500 mM. The cleaved sample was concentrated using Amicon Ultra system with a 3 kDa cutoff membrane (Millipore) and was then loaded onto a 24 ml Superdex 75 (GE Healthcare) size exclusion chromatography column equilibrated with GF buffer (20 mM Tris-HCl pH 8, 100 mM NaCl) with AKTA FPLC system. Peak fractions containing NTA visualized by SDS-PAGE with Coomassie staining were pooled, concentrated and flash-frozen in GF buffer and stored at -80 °C. ^{15}N labeled protein for NMR studies was purified following the same protocol, except that size exclusion chromatography was done in NMR buffer (20mM sodium phosphate buffer pH 7.5 and 100mM NaCl) and 5.5 % of D_2O was added to the final purified sample.

9. Gel-filtration chromatography and multi-angle light-scattering

For molar-mass determination of the MuB NTA 500 µl of purified sample at 0.6 mg ml^{-1} was fractionated by gel filtration on a Superdex 75 10/300 column equilibrated in GF buffer using an ÄKTA purifier at a flow rate of 0.5 ml min^{-1}. The eluted sample was characterized by in-line measurement of the refractive index and multi-angle

light scattering using Optilab T-rEX and DAWN 8+ instruments, respectively. The data were analyzed using the ASTRA 6 software to obtain the molar-mass (Wyatt, 1993).

10. Circular dichroism

A JASCO J-810 spectropolarimeter with a quartz cuvette of 0.1 cm path length was used for the circular dichroism experiments. Samples were prepared at a 10 µM concentration and the thermal melts were run at 1 °C min^{-1} monitoring the ellipticity signal at 222 nm, that reports changes in the protein secondary structure, with a bandwith of 2 nm and a data pitch of 0.2 °C. The thermal range of the experiments was from 20 °C to 99 °C and the recovery of the CD signal was over the 98% after thermal unfolding. The thermal denaturation curves from CD were fitted by a non-linear least squares fitting algorithm assuming the linear extrapolation method for a two-state unfolding without populated intermediate states (Santoro & Bolen, 1988).

11. NMR spectroscopy

NMR samples, both isotopic unlabeled and ^{15}N labeled, were prepared in 20 mM sodium phosphate buffer pH 7.5 and 100 mM NaCl, , and 5.5 % (v/v) D_2O at 500 µM MuB-NTA concentration. NMR spectra were recorded at 25 °C on a Bruker Avance 700 MHz spectrometer equipped with a triple resonance probe. The chemical shift assignments were based on a two dimensional (2D) TOCSY (Griesinger, Otting, Wüthrich, & Ernst, 1988) spectrum recorded with 20 ms mixing time and a three dimensional (3D) HNHA (Vuister & Bax, 1994) to assign the intraresidue alpha protons ($^1H\alpha$). A 2D NOESY (Jeener, Meier, Bachmann, & Ernst, 1979) and a 3D ^{15}N-NOESY-HSQC (Cavanagh, Fairbrother, Palmer, Skelton, & Rance, 2010) spectra (150 mixing time) were used together with a 2D TOCSY (100 ms mixing time) to establish the sequential connectivities and all ^1H side-chain assignments. The watergate pulse sequence was used for water suppression (Piotto, Saudek, & Sklenár, 1992). The carbon chemical shifts were assigned from a natural-abundance 2D ^{13}C-HSQC spectrum (Cavanagh, Fairbrother, Palmer, Skelton, & Rance, 2010) measured in D_2O.

All spectra were processed with NMRPipe (Delaglio, Grzesiek, Vuister, Zhu, Pfeifer, & Bax, 1995) and visualized and analyzed with the CcpNmr analysis suite (Vranken, et al., 2005). Overall, 97.6% of all the backbone amide and aliphatic side-chain proton resonances were assigned. The amide resonance of T32, N44, G45 and D46 were not observed probably due to conformational equilibrium and/or rapid exchange with the solvent. No resonances were identified as well for the $C_\alpha H_\alpha$ of N44 and D46 and for the $C_{\varepsilon 1} H_{\varepsilon 1}$ of H62. Two out of four arginine $N_\varepsilon H_\varepsilon$ were observed (R7, R56).

12. NOESY cross-peak assignment and structure calculation

3D structures were calculated with CYANA (Mumenthaler, Güntert, Braun, & Wüthrich, 1997; Güntert, 2003) on the basis of the sequence specific chemical shift assignments and NOE peak lists generated with CcpNmr analysis. Dihedral ϕ and ψ backbone torsion angle restraints were predicted with the program TALOS+ (Shen, Delaglio, Cornilescu, & Bax, 2009) and incorporated as well in the calculations. Seven cycles of automated NOE assignment and structure calculation were run, followed by a final structure calculation that uses only unambiguous assigned distance restraints. The structure calculation is started in each cycle from 100 conformers with random torsion angle values. The 20 conformers with the lowest final CYANA target function were retained for analysis and guide the NOE assignment in following next cycle. Assignments with an overall probability below 10 % in cycle 1 or 20 % in cycles 2-7 were discarded. The 20 conformers with the lowest final CYANA target function values were immersed in an 8 Å shell of explicit water molecules and subjected to restrained energy minimization against the AMBER force field (Cornell, et al., 1995) using the program OPALp (Luginbühl et al., 2000). CYANA was used to obtain all statistics (Table 3) on NOE assignments, conformational restraints, target function values, restraint violations and Ramachandran plots according to PROCHECK conventions (Laskowski et al., 1996). Conformational energies were calculated with OPALp. Structures were visualized and represented with MOLMOL (Koradi et al., 1996) and/or PyMOL (Schrödinger, 2010).

Results

1. Expression and purification of MuB protein

To tackle the structural and functional characterization of proteins the first, and perhaps the most important step, is to obtain large amounts of pure sample in a homogeneous conformational state. This posed an important challenge for MuB since, as already mentioned, the protein is exceedingly insoluble and exhibits a great propensity to aggregate (Chaconas et al., 1985). Additionally, MuB is highly sensitive to proteolysis, so it is difficult to overexpress and purify it in the absence of contaminating degradation products (Hung et al., 2000). Thus, a special protocol was designed and optimized for the purification of intact pure MuB in sufficient amounts to pursue structural studies.

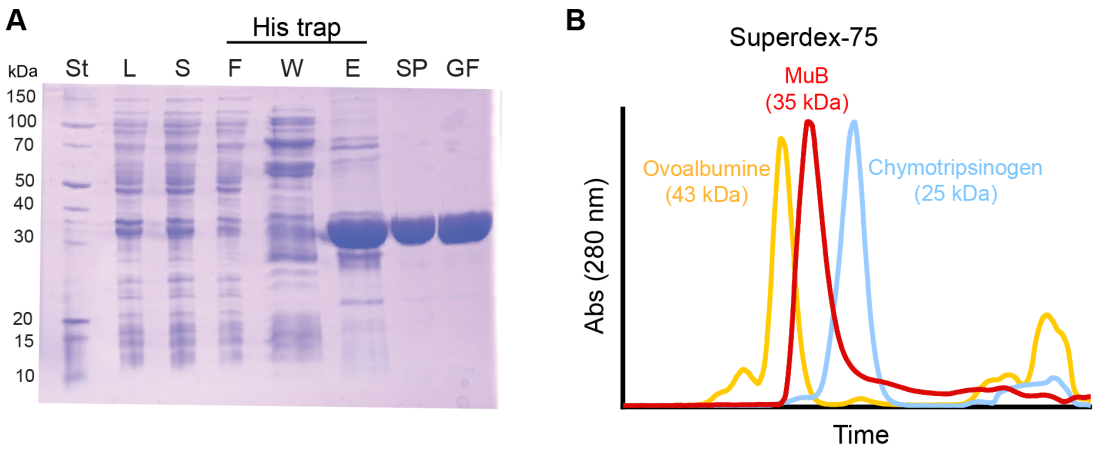


Figure 4. Purification of MuB. (A) Purity of the sample monitored by SDS-PAGE. St, molecular-mass standards; L, lysate; S, soluble fraction; F, W and E, flowtrough, wash and elution from His trap; SP, ion exchange column; GF, gel filtration. (B) Gel filtration elution profile of MuB compared to the elution peaks of two protein standard.

Recombinant MuB was expressed in *E.coli* by transforming the bacteria with the MuB gene (encoding residues 1 to 312) cloned in plasmid pET14b (Yamauchi, 1998), which appends a His_6-tag to the N-terminus of the protein. The construct includes a proteolytic cleavage site that allows removal of the affinity tag by digestion with thrombine. During this work, we decided to clone MuB in the vector pOPIN-B, which fuses an N-terminal His_6-tag cleavable by PreScission protease site. PreScission protease offers some advantages over thrombin, since it is more specific than

thrombin, works efficiently at high-salt concentrations and at reduced temperatures, and can be inexpensively produced in the laboratory.

MuB expression was induced either by the addition of IPTG. Following cell disruption and lysate clarification by centrifugation, the protein was purified in a three-step procedure involving immobilized metal ion affinity chromatography, anion-exchange chromatography and size-exclusion chromatography. A band of the overexpressed protein was clearly observed in the supernatant of the lysated cells and its purity after every step was monitored by SDS-PAGE (Figure 4 A). To minimize protein degradation, all purification steps were carried out at 4 °C and completed within a maximum of 1-2 days. In addition, to reduce the activity of metallo-proteases, the eluted protein from the metal affinity column was directly collected in fraction tubes containing EDTA. Following this step, the His_6-tag was cleaved off by the addition of thrombin or PreScission, prior to the ion-exchange step. MuB requires high concentrations of salt to be soluble. Thus, the initial purification steps are done with 1M NaCl in the buffer. However, in order to bind the protein to the SP-column, the ionic strength was reduced and compensated with the addition of 20% glycerol. This was carried out by dialysis during the incubation with the protease. For reason that would become apparently clear later on (see discussion), a significant fraction (up to 40 %) of the protein precipitated during this step. Following SP anion-exchange column, salt concentration in the buffer was increased to 1M, and the protein could be easily concentrated to 9 mg/ml. At this stage, 20 mg of protein were obtained per liter of culture. The protein was flashed-frozen in liquid nitrogen and stored at -80 °C or immediately passed on a next step of the purification if the tag was not cleaved. As a final purification step, the protein was passed through a size-exclusion chromatography. The protein elutes a single symmetric peak at approximately 13 mL, which according to the calibration of the column, corresponds to a monomer (Figure 4 B). Therefore, we managed to produce the large amounts of MuB of high purity and in homogenous oligomeric state, as required for structural studies.

We used a high-throughput approach to attempt the crystallization of MuB. The robotic platform for crystallization allowed us to test several thousands of commercial crystallization kits, at different protein concentrations and in buffers with different salts and glycerol concentrations. We tried sitting and hanging drop

crystallization, batch crystallization, and capillary free interface diffusion techniques. Moreover, the sample was taken to an EMBO training workshop, where it was subjected to more extensive crystallization trials. However, despite all of our efforts, MuB proved to be refractory to all crystallization trials.

One plausible reason for the impossibility of growing protein crystals could be that proteolyisis occurs during the crystallization trials, introducing undesired heterogeneity in the population of molecules to be crystallized. Indeed, the purified protein undergoes severe degradation within 1-2 days storage at 4 °C (Figure 5 D). These degradation products are weakly visible as two bands of 10 and 25 kDa, judged from SDS-PAGE, in the purified sample (Figure 5 A). Limited proteolysis assays using trypsin indicated that the two contaminating bands are MuB degradation products (Figure 5 B). N-terminal sequencing of the two degradation bands allowed us to identify precisely the fracture point. The spontaneous cleavage of purified MuB occurs between residues I231 and N232 (Figure 5 C). This proteolytic site is near the fragile sites previously described when MuB was treated with trypsin and chymotrypsin, although in these studies a precise sequence of the cleavage could not be defined.

The group of G. Chaconas had demonstrated that the MuB C-terminal fragment is exceedingly insoluble, although they managed to determine the NMR structure of the protein domain (Hung et al., 2000). This required purifying the protein in denaturing conditions and refolding in high salt concentrations. Based on our results we designed a new construct encoding for the N-terminal (1-231; MuB-ΔC) fragment of MuB. The protein was fused to a His_6-tag and purified following a similar protocol to full-length MuB. MuB-ΔC showed lower solubility than full-length MuB and crystallization attempts were also fruitless.

Expecting that binding of the substrates, ATP and DNA could stabilize MuB and induce some rigidity that could favor crystallization, we tested conditions with different ATP concentrations and in the presence or absence of short oligonucleotides. However, as previously observed by S. Ramón-Maiques, the addition of ATP, even at reduced concentrations caused heavy precipitation and hampered any crystallization attempts.

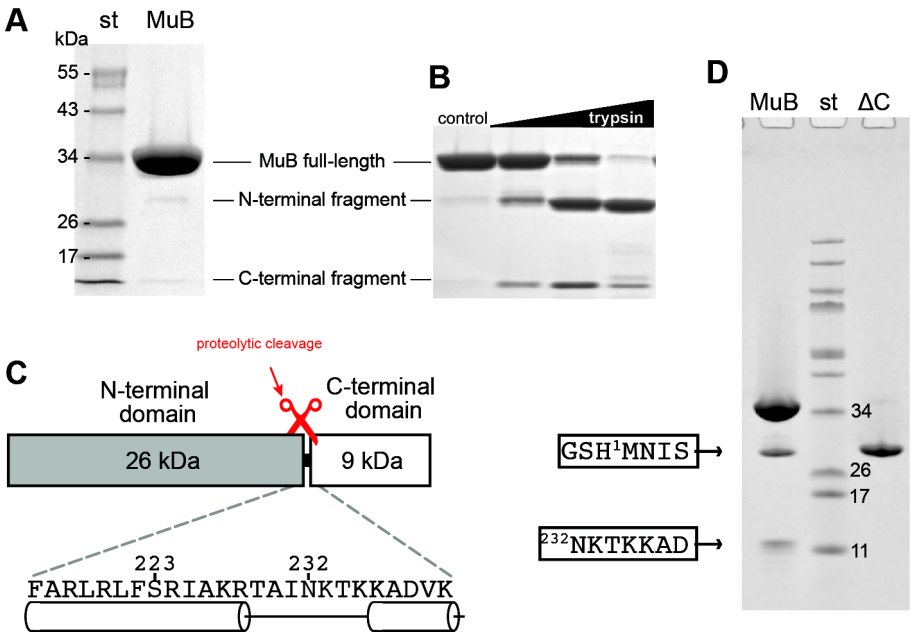


Figure 5. MuB degrades into two fragments. (A) SDS-Page showing purified MuB with two small contaminants. (B) Limited-proteolysis assays indicate that these bands correspond to the cleavage of MuB into its N- and C-domains. (C) Scheme of MuB protein, indicating the two protein domains and highlighting the sequence of the linker region. The cylinders underneath the sequence represent the predicted secondary elements. (D) N-terminal sequencing of the fragments corroborates that these contaminant bands are the degradation products and identify the fracture point.

2. Visualization of ATP-induced MuB filaments in the absence and presence of DNA

Intrigued by the strong tendency of MuB to precipitate in the presence of ATP, and given our lack of success with the crystallization trials, we decided to further study the MuB protein aggregates by electron microscopy (EM) (Figure 6).

Using negative staining EM we tested the formation of MuB filaments under different conditions. As previously observed by S. Ramón-Maiques, upon incubation with ATP, MuB forms filaments (Figure 6 A). Polymerization was fast, with filaments readily visible on grids prepared only a few seconds after mixing MuA and ATP. The filaments had a uniform width of ~150 Å and varied in length, typically between 200 and 600 Å, and they exhibited a stain penetrable axial channel. Ring-shaped particles

with a diameter of 150 Å and a hollow center were also seen, which we think represent top views of short filaments.

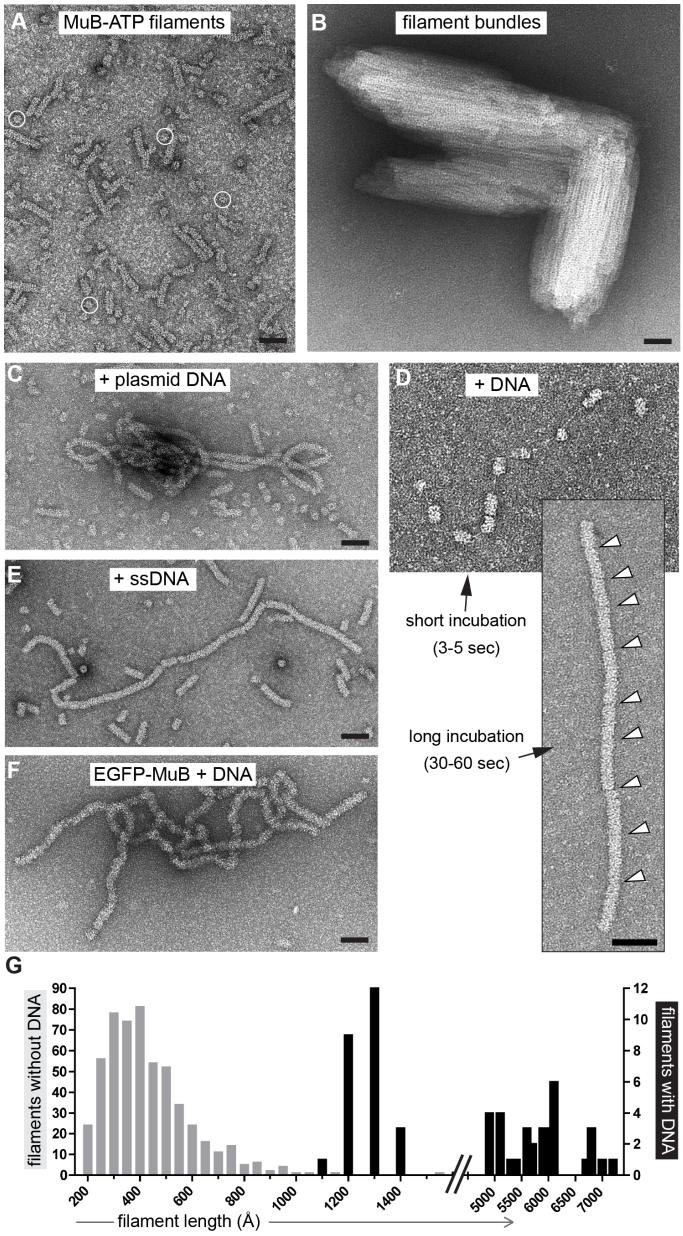


Figure 6. MuB filaments viewed by negative-staining EM. (A) ATP-induced MuB filaments. Ring-shaped particles, representing axial views of very short filaments, are encircled. (B) MuB filaments aggregate forming bundles. (C) MuB filaments on a plasmid DNA. (D) MuB filaments covering a linear double stranded DNA. At short times, the DNA is partially covered by MuB filaments. As the filaments grow the DNA becomes fully covered. Arrowheads indicate the discontinuities along the filament. (E) Filaments also form on single stranded DNA. (F) Filaments of EGFP-MuB. (G) Length of DNA(−) (gray) and DNA(+) filaments (black). The distribution of the filament lengths agrees with the expected sizes of the DNA fragments: 0.34 kb (1,200 Å), 1.5 kb (5,100 Å), 1.7 kb (5,800 Å), and 1.8 kb (6,100 Å). Scale bars, 500 Å.

Filament formation required the addition of Mg^{2+} in the buffer and was not detected without ATP or with ADP. However, similar filaments were observed in the presence of ATPγS, a non-hydrolyzable nucleotide analogue, confirming that ATP binding but not hydrolysis is needed for MuB to polymerize (Adzuma and Mizuuchi, 1991). Interestingly, by increasing the incubation time with the nucleotide, we observed that the filaments aggregate into large bundles. (Figure 6 B). This large filament bundles might explain the low solubility of MuB upon addition of ATP.

To test the effect of DNA on MuB filament formation, we incubated the protein with a plasmid DNA for a few seconds in the presence of ATP. Strikingly, the EM images showed long and curved filaments, as if the plasmid was fully coated by MuB (Figure 6 C). To facilitate the visualization of the MuB-ATP-DNA complex, the plasmid was digested with restriction enzymes to generate fragments of defined sizes ranging between 0.34 and 1.8 kb. At short incubation times (3-10 seconds), short tracts of MuB filament could be seen distributed at irregular intervals along the DNA molecules (Figure 6 D), indicating that filament assembly had initiated at multiple sites. Following longer incubations (~1 min), fully coated DNA molecules were observed (Figure 6 E). These long MuB filaments are interrupted with discontinuities, suggesting that the segments initiated at different nucleation sites had extended but could not annealed, perhaps because they have different polarities or by some other physical constraints. Otherwise, the filament morphology did not change when DNA was included, suggesting that the DNA is inside the filament

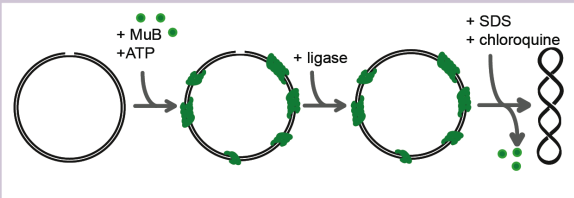
We also observed that MuB filaments form on single-stranded DNA. Although the DNA is not visible by negative staining, the long MuB filaments exhibit sharp kinks, as if the different MuB segments were threaded by the nucleic acid. In the course of the trials, we also found that, by increasing the salt concentration from 0.15 to 0.3 M, MuB-ATP polymerization became DNA-dependent; only long DNA-bound filaments were observed. This allowed us to specifically prepare MuB filaments assembled on the DNA.

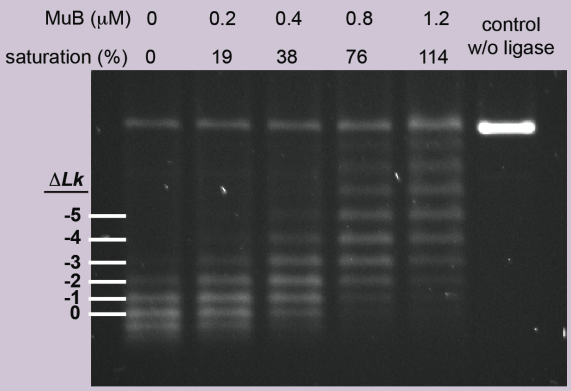
Clearly, the ATP-induced filaments are longer in the presence of DNA than without DNA [hereafter named DNA(+) and DNA(-) filaments]. The median length of the DNA(+) filaments matched the expected length of the B-form DNA molecules (Figure 6F), indicating that MuB binding does not deform the DNA B-form structure. This

agrees with previous single-molecule assays where the addition of EGFP-MuB to an immobilized phage lambda DNA did not significantly modified the length of the DNA (Greene & Mizuuchi, 2004). Our results, also confirmed unpublished experiments from the group of Kiyoshi Mizuuchi, in which by following changes in topoisomer distribution, they proved that DNA topology is not altered by the binding of MuB (Box 2).

Box 2

MuB filament binding does not alter DNA structure or topology





The group of **Kiyoshi Mizuuchi** tested if MuB filament affect DNA conformation by checking whether linking number of relaxed circular DNA changes when bound by MuB. Nicked circular pBR322 DNA was converted to closed circular form in the presence of different concentrations of MuB and ATP, deproteinized and analyzed by agarose gel electrophoresis in the presence of chloroquine to examine the topoisomer distribution. The topoisomer distribution broadened in the presence of higher concentrations of MuB, suggesting increased torsional flexibility within the MuB filament. This is unlikely to be caused by increased bending flexibility, considering that the MuB-coated filament appeared stiffer than naked DNA by EM. The center of topoisomer distribution shifted toward reduced linking numbers as more MuB was added to the DNA solution. Combined with the information from the binding titration experiments (see Box 1), it was estimated that DNA within MuB-bound filament is only approximately 1.2 % under-wound compared to the relaxed B-form DNA. Therefore the DNA conformation within MuB filament appears to be close to the B-form structure.

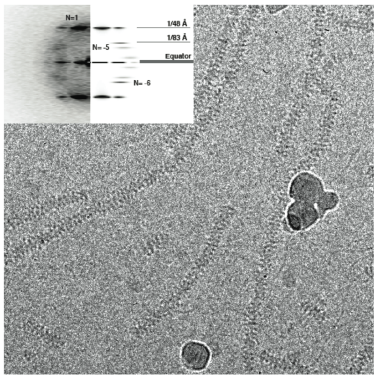
To get further insight into MuB filament, we collaborated with Naoko Mizuno, to do single-particle cryo-EM on unstained MuB filaments. MuB-ATP filaments were vitrified and imaged by cryo-EM (Box 3). Power spectra of relatively straight filament were indicative of helical symmetry. The filament was a single-start helix with an axial spacing of 48 Å. Following image classification and averaging, Naoko obtained 3D-reconstrucions of MuB filaments. The model shows a solenoid with a pitch of 48 Å and ~5.4 subunits per turn, and an axial channel 35-40 Å across. DNA(+) and DNA(-) MuB filaments show similar helical parameters, and although then inner channel is wide enough to accommodate a double stranded DNA, the DNA(+) filaments did not

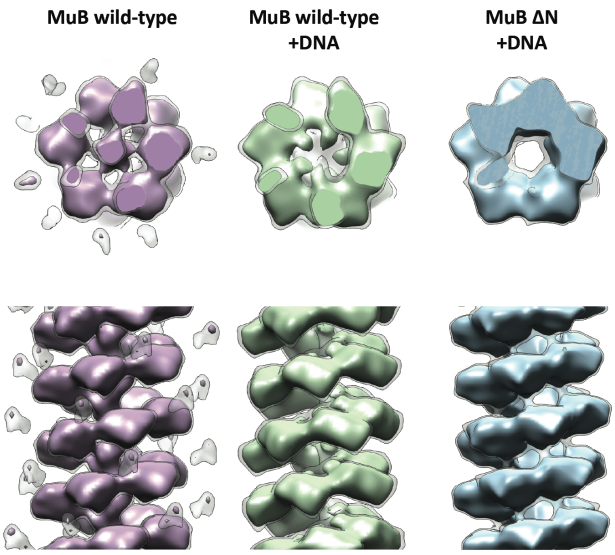
show clear additional density for the DNA. This is most likely because the DNA density is blurred out during the reconstruction procedure.

Box 3

3D reconstruction of MuB filaments by cryo-EM

3D structures of MuB filaments were determined by **Naoko Mizuno**. MuB filaments were vitrified and imaged by cryo-EM.





Power spectra showed a strong layer line at axial spacing of 48 $Å^{-1}$, indicating that the MuB filaments observe helical symmetry. The filaments are right-handed as observed from metal-shadowed specimens (not shown).
The images of filaments were dissected into short segments, and this data set was subjected to reference-free classification, averaging and 3D reconstruction using IHRSR methods. 3D reconstructions show a solenoid structure (single-start helix) with ~66.4º twist and ~9 Å rise per subunit, a pitch of 48 Å and an axial channel of ~35-40 Å.
3D reconstructions of filaments with DNA show that the density for the DNA molecule is weak. This is most likely due to the helical symmetry mismatch between the MuB filament and the DNA, which causes the DNA to be delocalized in the particle set and is averaged out during the reconstruction process. The reconstruction with the most sharply defined features showed a small density, protruding into the axial channel, which probably correspond to DNA binding loop-1 of MuB (see more below).
Images of MuB-ΔN filaments were also recorded in an effort to localize the N-domain. The reconstructed model of the MuB-ΔN filaments with DNA was similar to that of MuBwt. However, discrete density that could be assigned to the N-terminal appendage was not detected. As MuB-ΔN filaments in the absence of the DNA were shorter, informative reconstructions could not be preformed. The results indicate that the N-domain is flexibly attached to the AAA+ module. A similar situation has been described for the AAA+ protein unfoldase, ClpA.

3. MuB is predicted to be an AAA+ ATPase

The discovery of MuB filaments and the 3D reconstructions were very compelling but we lacked a mechanistic understanding of how ATP controls MuB polymerization and how the filaments interact with the DNA and target if for transposition.

To get insight into the nature of MuB, we used the protein structure prediction servers PHYRE (Kelley et al., 2009) and HHpred (Söding et al., 2005) to search for structural homologs of MuB. Although in the past no sequence similarity between

MuB and other proteins was reported, surprisingly, both search engines found a high similarity between MuB and a number of proteins. Most interestingly, all the proteins belong to the AAA+ ATPase superfamily. AAA+ ATPases are characterized by a bi-modular monomer structure that binds and hydrolyzes ATP, and typically oligomerizes into ring-like or helical assemblies. Although sequence identity is low (8-17 %) MuB residues 77 to 312 were predicted to have the secondary and tertiary structure characteristic of AAA+ ATPases (Table 2).

Table 2. Structural homologs of MuB obtained with Phyre (Kelley et al., 2009).

Protein (PDB ID)	Residues aligned	Sequence identity (%)	Confidence*
NtrC1 (1NY5)		**19**	**99.8**
PspF (2BJV)		**17**	**99.7**
ZraR (1OJL)		**18**	**99.7**
RFC (1SXJ)		**16**	**99.6**
DanA (3R8F)		**12**	**99.7**

* Confidence represents the probability (from 0 to 100) that the match between input sequence and Phyre template is a true homology.

In addition to finding far protein homologues, PHYRE calculates a 3D model of the protein of interest. The high confidence match (>95 %) calculated by PHYRE indicated that the overall fold of the MuB model is almost certainly correct and the central core of the model will tend to be accurate. These results suggested that MuB is formed by three distinct segments: a putative AAA+ ATPase module composed of (1) a central α/β domain (residues 77-231) and (2) a C-terminal helical bundle (residues 232-312), and (3) an N-terminal appendage (NTA) (residues 1-76) (Figure 7 A).

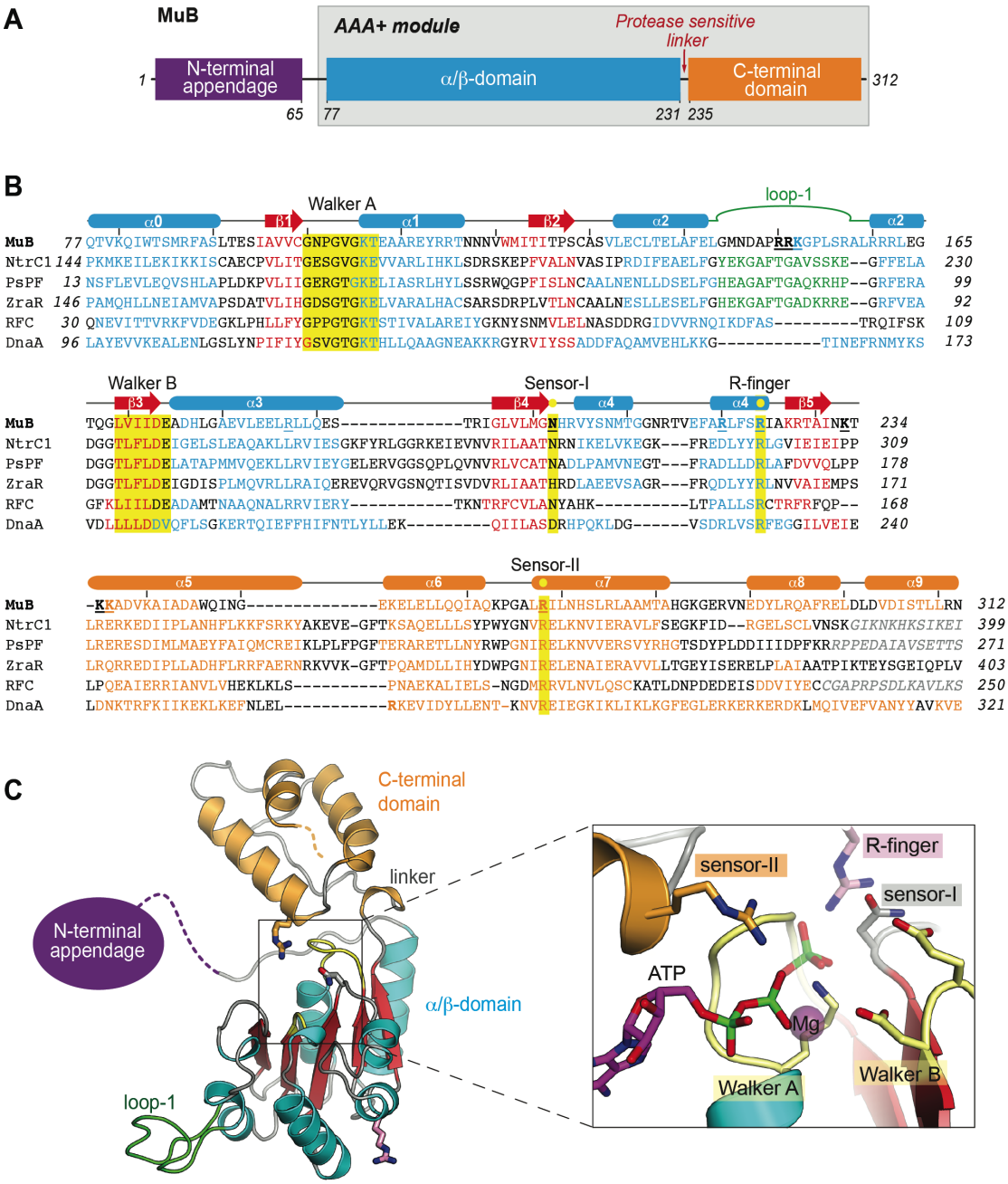


Figure 7. MuB is an AAA+ ATPase. (A) MuB is predicted to have three different segments: an N-terminal appendage (in violet) and an AAA+ module composed of a central α/β-domain (in cyan and red) and a C-terminal helical bundle (in orange). (B) The AAA+ module of MuB is aligned with the AAA+ modules of NtrC1 (PDB 1NY5), PsPF (PDB 2BJV), ZraR (1OJL), RFC (1SXJ), and DnaA (3R8F). Predicted MuB secondary structural elements match those observed in the structures of the aligned proteins. Conserved α-helixes (cylinders) and β-strands (arrows) are represented above the sequences. AAA+ elements are shown with a yellow background. Residues in gray were not observed in the crystal structures. Underscored MuB residues were mutated in this study. (C) Cartoon representation of MuB AAA+ domain model calculated by Phyre and detailed view of the active site of MuB based on NtrC1 and RFC structure.

Interestingly, the protease sensitive region that produces two MuB fragments is located between the predicted nucleotide binding α/β domain and the C-terminal helical bundle of the AAA+ module (Figure 7 A, B). This would explain that this linker region is more flexible and accessible to proteases. It also explains why, as already mentioned, MuB-ΔC (residues 1-231) had solubility problems. According to our prediction the C-terminal fragment is an integral part of the AAA+ module and therefore it is essential for the function of the AAA+ module.

To test the predicted architectural module of MuB, we produced a mutant consisting of only the putative AAA+ module by eliminating the predicted N-terminal appended sequence (MuB-ΔN; residues 65-312) (Figure 8). Contrary to the MuB-ΔC mutant, MuB-ΔN is well expressed, soluble and stable. Following a similar procedure as for MuBwt, MuB-ΔN was isolated and the final yield and purity was comparable to that obtain for the intact protein (Figure 8). MuB-ΔN elution from gel filtration is slightly retarded compared to full-length protein, indicating that the mutant is also a monomer in solution (data not shown).

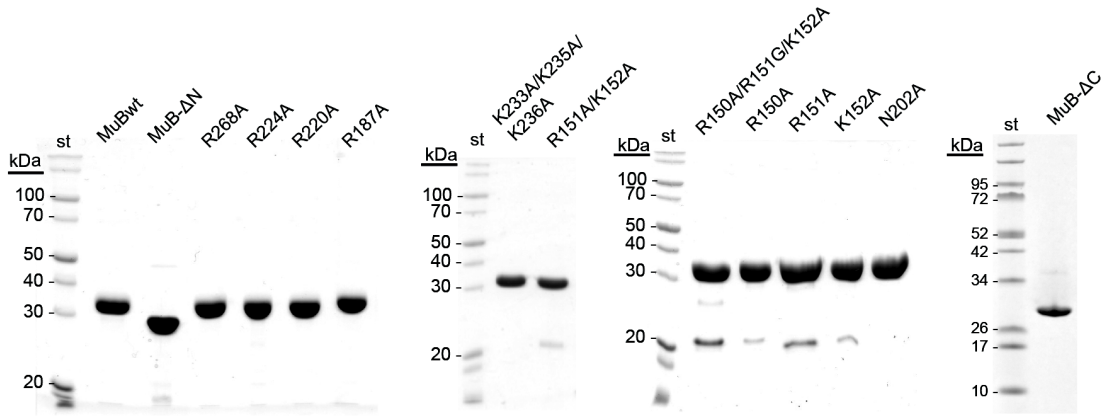


Figure 8. SDS-PAGE loaded with 5 µg of the purified samples of MuBwt and mutant forms. Some preparations had minor contamination of an ~20-kDa protein, which was identified as Catabolite Activator Protein (CAP) by sequence analysis.

Next, we studied if MuB-ΔN and MuB-ΔC had ATPase activity and if this activity responded to the addition of MuA and DNA in a similar manner as wild-type MuB. It is known that the ATP hydrolysis rate of MuB is inhibited by DNA and significantly stimulated by MuA in the presence of the DNA, but not by a truncated form of MuA

lacking the C-terminal 48 residues (MuA_{trun}) (Maxwell et al., 1987; Adzuma and Mizuuchi, 1991). To test this, we mixed radioactive ATP-α^{32}P with MuB in the presence or the absence of the DNA and/or MuA or MuA_{trun} and separated the ADP produced by MuB using thin layer chromatography (Figure 9 A). Our results showed that MuB-ΔN exhibits the ATPase activity at 50% of full-length MuB (k_{obs} = 3.2 min-1) (Figure 9 B), and similarly to MuBwt, the activity is activated by MuA and inhibited by DNA (Figure 9 C). Contrary, MuB-ΔC showed not detectable ATPase activity even at high protein concentrations (Figure 9 B).

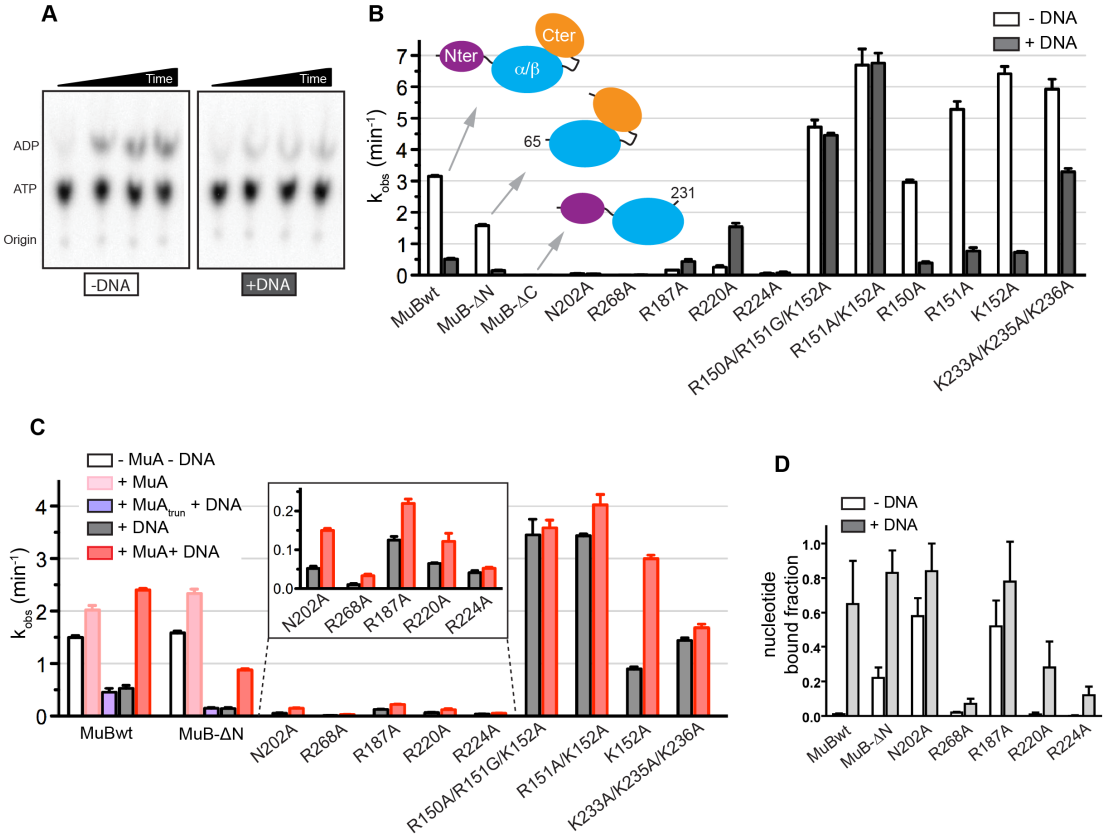


Figure 9. Site-directed mutagenesis and activity assays confirm that MuB is an AAA+ ATPase. (A) ADP produced by MuB in the presence and the absence of the DNA separated from ATP by thin layer chromatography. (B) ATPase activity of MuBwt and mutant forms in the presence and absence of DNA. (B) Effect of MuA and DNA on ATPase activity. (C) Nucleotide-binding assay by centrifugal gel filtration.

We also wanted to know if the MuB-ΔN binds ATP in the same way as wild-type protein. For this, we did nucleotide-binding assays in which we incubated protein with the radioactive ATP-α^{32}P and applied the mixture to a centrifugal gel filtration (Penefsky, 1977; Ramón-Maiques et al., 2002). In this assay, unbound nucleotides

get retained inside the column, while protein-ATP complex appears in effluent. MuB wild-type showed a very small amount of nucleotide-bound protein in the effluent, indicating that ATP dissociation is a fast event, or that during pre-incubation (10 s) and centrifugal passage (1 min), ATP is hydrolyzed (ATPase turnover is $t_{1/2}$= 30 sec) and ADP dissociates and is retained in the column (Figure 9 D). Indeed, the nucleotide-bound fraction increased to ~60% when DNA was added to the mixture, in agreement with the inhibitory effect of DNA on ATPase activity. Similarly, 20 % of MuB-ΔN eluted from the column bound to nucleotide, likely due its reduced ATPase rate, and the nucleotide-bound fraction increased nearly 4-fold in presence of DNA.

We also tested the ability of MuB-ΔN to polymerize by negative staining EM. We observed that the truncated protein forms filaments upon ATP addition and as shown for the wild-type protein, they become longer in the presence of the DNA (Figure 10).

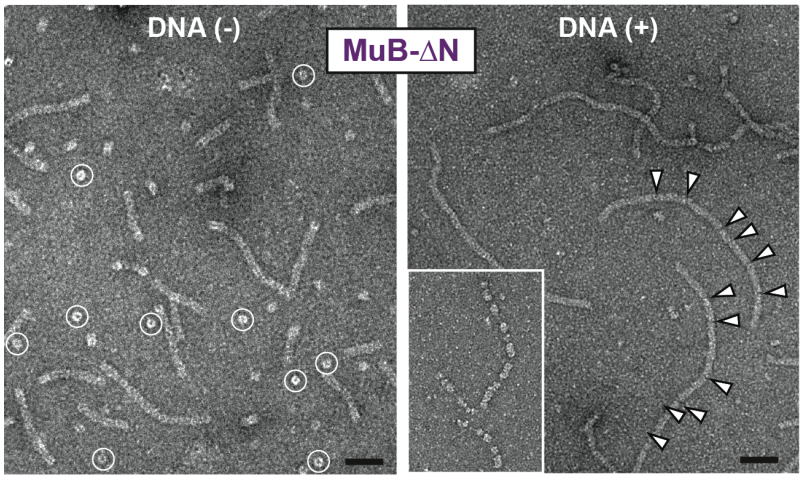


Figure 10. MuB mutant filament formation viewed by negative-staining EM. (A) MuB-ΔN filaments formed in absence and presence of DNA. In the DNA(−) sample, ring-shaped particles, representing axial views of filaments, are encircled. The white triangles point to filament discontinuities. (Inset) DNA molecule partially covered by MuB-ΔN filaments. (Scale bar, 500 Å.)

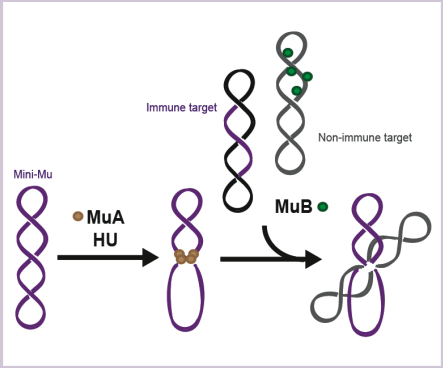
MuB-ΔN filaments were also used for cryo-EM and N. Mizuno obtained a 3D reconstruction of the filament bound to DNA (Box 3). The model showed similar helical parameters than those of MuBwt. Comparison of the MuBwt and MuB-ΔN models did not show an extra density that could be attributed to the N-terminal

appendage. Thus, it indicates that the N-terminal part of the protein is flexibly attached to the AAA+ module and its density is averaged out during reconstruction process.

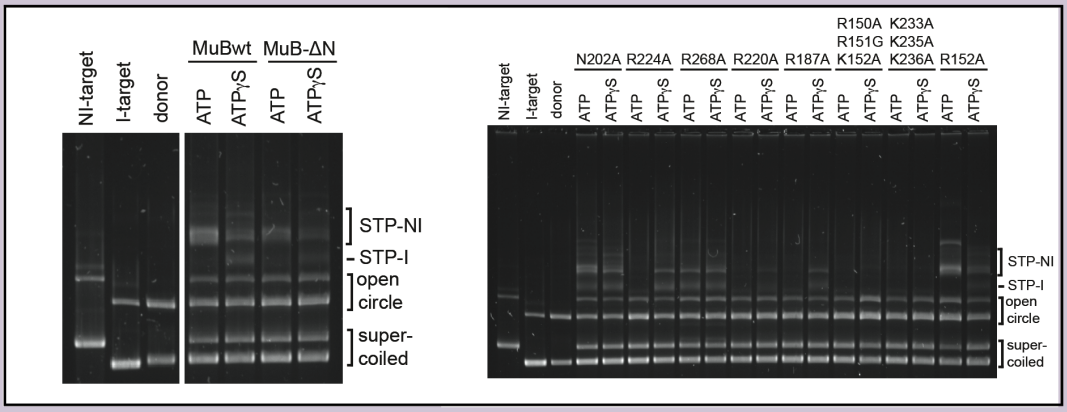
To further test the functionality of MuB-ΔN Kiyoshi Mizuuchi tested the ability of MuB-ΔN (and of a battery of other MuB mutants described below) to stimulate inter-molecular transposition reaction in a cell-free reaction system. These experiments prove that MuB-ΔN is able to stimulate inter-molecular transposition reaction and discriminate against "immune" target DNA in the same way as the wild-type protein (Box 4).

Box 4

Strand-transfer stimulation activity of MuB



Kiyoshi Mizuuchi tested the ability of the MuB mutant proteins to stimulate strand transfer reaction. Strand-transfer products (STP) made in reactions that contained both immune (I) and nonimmune (NI) targets and either ATP or ATPγS were examined. Presence of both STP-I and STP-NI in the reaction with ATP, as in reactions in the presence of ATPγS, indicates strand-transfer stimulation without discrimination against immune target. Presence of STP-NI but no (or much less) STP-I in the reaction with ATP indicates strand-transfer stimulation with discrimination against immune target. The results show that MuB ΔN and K152A stimulate intermolecular strand transfer and exhibits target immunity. N202A stimulate intermolecular strand transfer and exhibits very weak target immunity. R224A and R187A stimulate intermolecular strand transfer only in the presence of ATPγS. R268A stimulates intermolecular strand transfer and exhibits no target immunity. R220A stimulates intermolecular strand transfer very weakly only in the presence of ATPγS. R150A/R151G/K152A and K233A/K235A/K236A do not stimulates inter-molecular strand transfer.



In summary, these results confirm that MuB consists of an AAA+ module that is needed and sufficient for ATP binding/hydrolysis and for filament formation. This module is preceded by an N-terminal appendage, which can be removed without severely impeding any of the functions of MuB.

4. Identification of AAA+ ATPase characteristic elements: sensor I, sensor II and arginine-finger

If the prediction of MuB protein structure is correct, it should allow us to identify the residues most likely to be involved in the function of MuB. In addition to the Walker A and Walker B motifs commonly present in many nucleotide binding proteins, AAA+ ATPases posses characteristic residues involved in ATP binding and hydrolysis. The importance of the Walker A and Walker B motifs of MuB, was demonstrated previously by the group of T. Baker (Yamauchi et al., 1998), who showed that mutation K106A (Walker A) and E174Q or D176H (Walker B) strongly diminished ATPase activity.

Our similarity analysis predicts the location of these functionally important residues in MuB: N202 was predicted to be sensor I, R268 sensor II and R224 to be R-finger (Figure 7 B, C). To test this, we mutated these residues to alanine and produced single-point mutants that were expressed and purified following the protocol described for MuB-wt (Figure 8). Next, we tested their ATPase activities and compared them with the activity of the wild type protein (Figure 9 B). N202A and R224A decreased the ATPase activity ~60-fold, whereas the activity of R268A was practically undetectable. Two other arginines, R187 and R220, were predicted to be at the same protein interface as the R-finger, and mutations of these residues to alanine reduced the ATPase rate by 20- and 12-fold, respectively. The importance of these arginines will be discussed later on.

The ATPase activities of all the mutants increased by addition of MuA in the presence of the DNA, but still they remained low compared to wild-type (Figure 9 C).

To distinguish between defects in ATP hydrolysis versus ATP binding, we measured the ability of these mutants to form a complex with ATP or ADP (Figure 9D). Using

nucleotide-biding assay, we would expect that mutants that bind but cannot hydrolyze ATP would show a larger nucleotide-bound fraction in the effluent. This was the case of N202A and R187A, for which nearly 60% of the molecules are bound to nucleotide. Contrary, the nucleotide-bound fraction of R220A and R268A was less than 5%, indicating that these mutants are impaired to bind ATP. Interestingly, the addition of DNA also increased ATP binding in the mutants, a phenomenon that is not explained by the inhibitory effect of DNA on ATP hydrolysis, since the mutants are nearly inactive, and that must respond to an alternative mechanism.

We also tested the ability of the MuB mutants to forms filaments by negative staining EM (Figure 11). MuB N202A, which can bind but not hydrolyze ATP, forms filaments like those of MuB-wt. On the other hand, for those mutants with defects in ATP binding (R268A, R220A and R224A) polymerization was strongly impaired, and in absence of DNA we only observed ring-shaped particles. However, when polymerization was induced in the presence of DNA, all the mutants except for one, R187A, proved capable of polymerizing on the DNA, forming the easily detectable long nucleoprotein filaments. Giving the apparent difficulty of these mutants to polymerize on their own, this strongly suggests that DNA acts as a scaffold that facilitates ATP binding and protein-protein interactions. R187A on the other hand, did not form regular filaments despite its capacity to bind ATP. This suggests that R187 may contribute to stabilizing the MuB polymerization interface.

Strand transfer stimulation assays carried out by Kiyoshi Mizuuchi with the provided MuB single-point mutants proved that the mutations also alter the capacity of MuB to confer target immunity (Box 4). Although N202A, R224A and R268A mutants can stimulate the intermolecular strand transfer, and thus can allosterically activate MuA, they lack or exhibit very weak target immunity. This further corroborates the central role of these residues for MuB function and confirms the important role of filament formation to confer target immunity.

Overall, these results confirm the predicted functional role of N202, R268 and R224 as the sensor I, sensor II and arginine finger, respectively, and thus, authenticate MuB as a previously unidentified member of the AAA+ ATPases superfamily of proteins.

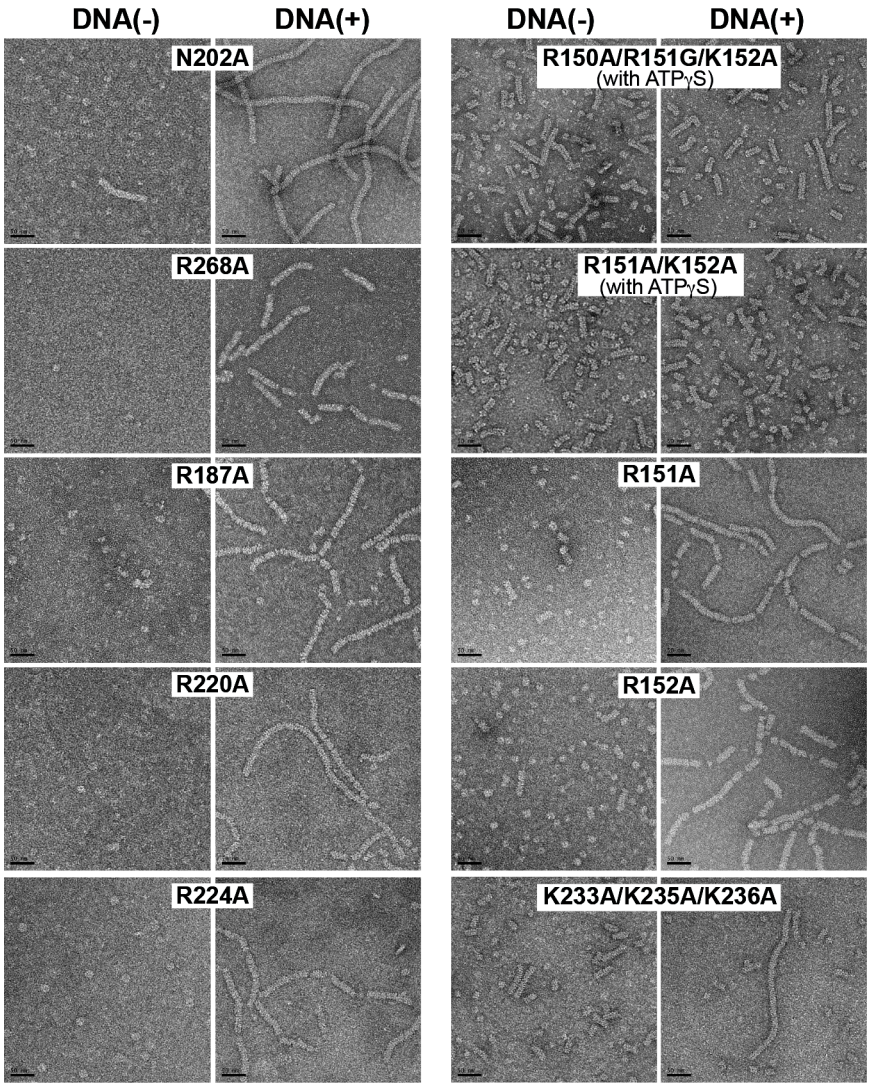


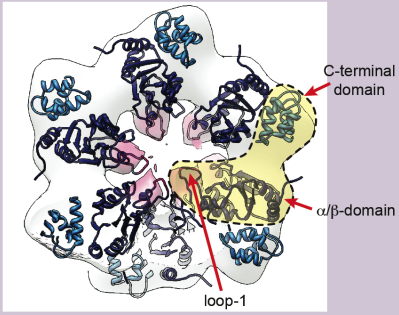
Figure 11. MuB mutant filament formation viewed by negative-staining EM. MuB mutants were incubated with ATP (or ATPγS when indicated) in presence or absence of DNA. (Scale bar, 500 Å.)

5. Identification of MuB DNA binding loop

Reassured by the biochemical identification of key AAA+ elements, N. Mizuno attempted the molecular docking of an AAA+ module into the cryo-EM density. The modeling experiment showed that the motif fit quite well into the EM envelope (Box 5). Then, we asked what elements are involved in the interaction with DNA. A key element that is conserved in a large number of AAA+ proteins, particularly in those showing highest similarity with MuB (NtrC1, PsPF and ZraA), is a loop (loop-1) that

protrudes from helix α2 and is involved in substrate (protein or DNA) recognition. In MuB, an eight-residue loop is predicted to insert in the middle of helix α2 (Figure 7 B, C).

Box 5



Molecular docking of an AAA+ module

Naoko Mizuno used the AAA+ module of NtrC1 (PDB ID 1NY5) as model for fitting in the EM envelopes of MuB reconstructions. The AAA+ module was found to fit quite well into the cryo-EM density. This modeling experiment implies that the quasi-globular unit of density repeated along the strand consists of the α/β-domain of one subunit and the C-terminal domain of the adjacent subunit. One of the reconstructions of MuBwt with DNA nicely accommodates the density of DNA binding loop-1 facing the axial channel.

Interestingly, one of the reconstructions of MuBwt with DNA shows a density protruding into the axial channel that nicely fits the loop-1 from NtrC1 (Box 3).

We made the triple mutation R150A/R151G/K152A and the double mutation R151A/K152A in this putative DNA-binding loop (Figure 8). These mutants have ATPase activities ¬~1.5- and 2-fold higher than MuBwt, but importantly, the activity is unaffcted by DNA (Figure 9 B). The ATPase activity does not respond significantly to MuA either, indicating that MuB-DNA binding is required for the ATPase to be stimulated (Figure 9 C). To dissect the contribution of each of the three residues to the higher activity and/or to the recognition of DNA, we individually mutated the three residues to alanine (Figure 8). R150A, R151A and K152A single mutants exhibited activities equal, 1.7-, and 2-fold higher than MuB-wt, respectively, but unlike the multiple mutants, the activity was inhibited by the addition of DNA, and K152A was stimulated by MuA (Figure 9 B, C). These results indicate that both R151 and K152 are involved in the interaction with DNA. Consistent with the above, the triple mutant failed to stimulate strand-transfer reaction, whereas the K152A mutant did not stimulate strand transfer and discriminate against immune target (Box 4).

In agreement with the lack of effect by DNA on the ATPase activity the length of the filaments formed by the multiple loop-1 mutants did not change upon addition of DNA (Figure 11), indicating that the mutants do not bind DNA. In contrast, the single

mutants R150A, R151A and K152A assembles into longer filaments on the DNA similarly to MuBwt.

6. Mutations in the linker abolish MuA stimulation

The proteolytic C-terminal fragment of MuB was previously proposed to bind DNA nonspecifically and to interact with MuA through a patch of three lysines, K233, K235 and K236 (Hung et al, 2000; Coros et al, 2003). We recognized the C-terminal fragment as an integral part of the AAA+ module, which is connected to the α/β-domain by a linker that harbors the three lysines. Because the C-terminal domain and the linker invariably locate at the periphery of the AAA+ assemblies, and in the MuB filament reconstructions they do not face the internal filament channel where DNA binds, we reexamined the role of these lysines in DNA recognition. We made a triple linker mutant K233A/K235A/K236A (Figure 8). This linker mutant can polymerize on the DNA (Figure 11) and its ATPase activity is inhibited by the addition of DNA (Figure 9 B). Therefore, we concluded that this patch of positively charged residues, and therefore the linker region, is not involved in DNA recognition. Interestingly, other AAA+ members (e.g. NtrC1) also present a linker with positively charged residues that undergoes conformational changes associated with the ATPase cycle (Rappas et al., 2006). Thus, an attractive possibility is that the ATPase cycle could be controlled by conformational changes at this linker region. The fact that the triple linker mutant, like the loop-1 mutants, enhances ATPase activity supports this notion (Figure 9 B). Because the ATPase activity of the K233A/K235A/K236A mutant and the DNA-bound filaments do not respond to MuA (Figure 9 C) and fail to stimulate strand-transfer reaction (Box 4), these lysine residues most likely interact with MuA to trigger ATP hydrolysis.

7. The N-terminal appendage of MuB promotes filament association

During all this characterization of the protein, the N-terminal part of MuB (residues 1-72), were appended to the AAA+ module under scrutiny but its presence did not

significantly influenced any of the activities being tested. In our interest to fully understand MuB we set to characterize this apparently useless appendage.

As briefly mentioned before, early observations showed that MuB-ATP filaments were "sticky" and exhibited a strong tendency to clump together into large filament bundles (Figure 6 B). This forces us to prepare the EM grids shortly (3-5 sec) after adding the ATP to the protein, since longer incubations (10-30 s) resulted in clumsy specimens were filaments could not be pictured nicely separated for further image processing and 3D reconstruction. This phenomenon was a technical problem that luckily could be avoided by shortening the polymerization time, and was considered an artifact that did not merit further attention. However, to our surprise, the filaments formed by MuB-ΔN did not show this "stickiness", and remained as independent filaments even after 1 h incubations (data not shown). This differential behavior was confirmed by testing bundle formation of MuBwt and MuB-ΔN at varying time, salt and glycerol concentrations, pH, etc. This indicated that the N-terminal appendage was involved in filament-filament interactions, and suggested that perhaps the observed tendency of the filaments to stick to each other is a relevant characteristic of MuB.

8. The N-terminal appendage of MuB is a well-folded globular domain

To better understand the role in filament interactions, we set to determine the three dimensional structure of MuB N-terminal appendage. Primary sequence analysis with PHYRE predicted with high probability that the NTA folds into four α-helices with short linkers connecting them. Despite low sequence identity (8-16 %) the search engine suggested with 100% estimated precision that NTA is homologous to proteins of the superfamily of λ repressor-like DNA binding domains.

Two different constructs encompassing MuB residues 1-63 or 1-72 were produced in *E. coli*. The proteins, fused to a cleavable N-terminal His_6-tag, were purified in a three-step purification (Figure 12 A). Although both protein constructs could be obtained with high purity and in good yields (5 mg per liter of culture), we focused our efforts in construct 1-63 since it was the shortest and, a priori, the most likely to

crystallize since the C-terminal 9 residues connecting with the AAA+ module were predicted to be disordered.

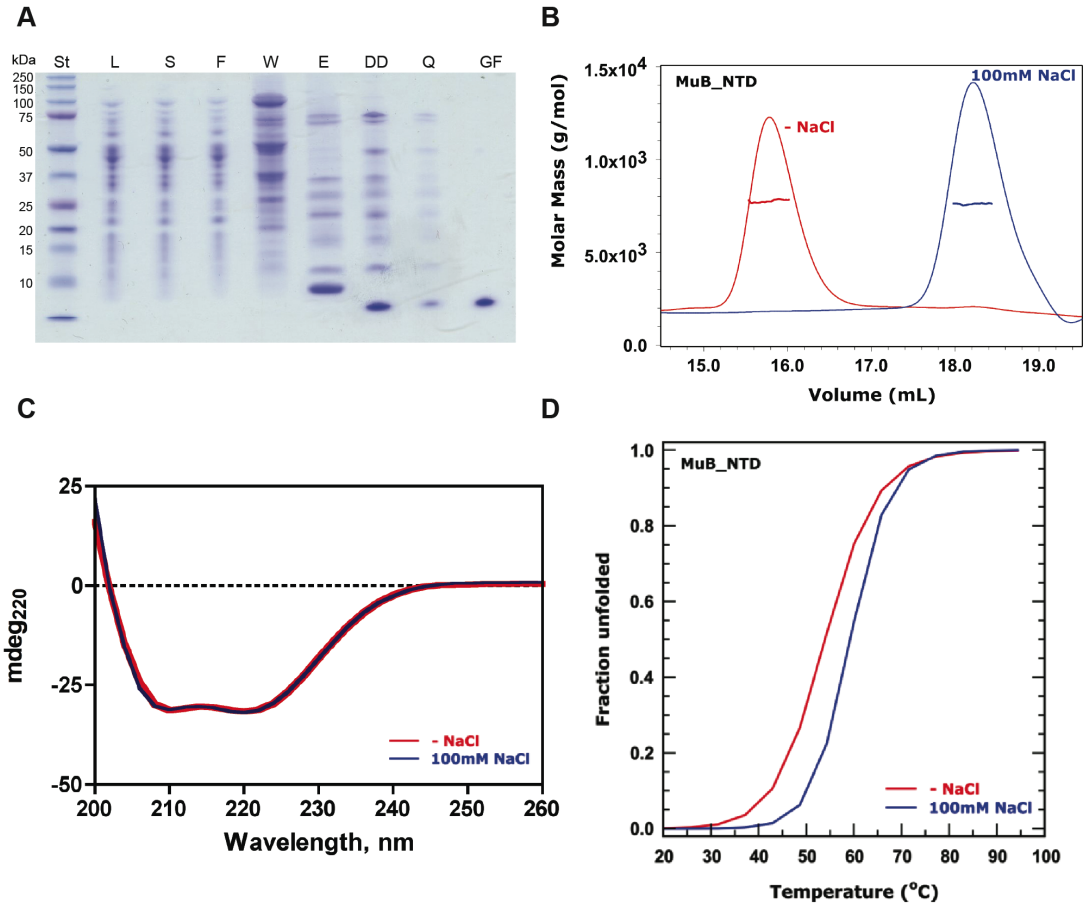


Figure 12. Purification and biophysical characterization of MuB N-terminal appendage. (A) Purity of the sample monitored by SDS-PAGE. St, molecular-mass standards; L, lysate; S, soluble fraction; F, W and E, flowtrough, wash and elution from His trap; DD dialyzed and digested sample, Q, ion exchange column; GF, gel filtration. (B) Gel filtration coupled with light scattering. (C) Circular dicroism experiments. (D) Thermal unfolding experiments. (B), (C) and (D) experiments were all done in the presence (100 mM NaCl) (blue) and the absence of salt (red).

Size-exclusion chromatography coupled to multi angle light scattering (MALS) analysis showed a single monodisperse peak with a calculated mass of 7.4 kDa (Figure 12 B), proving that the NTA is a monomer in solution. The protein is stable in Tris and phosphate buffers and the decrease in the elution volume from the size-exclusion chromatography in the absence of additional salt in the buffer, suggest a more globular structure in the presence of salt. Circular dichroism and thermal unfolding experiments confirmed this observation. The circular dichroism both in presence of salt (100mM NaCl) and with no salt in the buffer is consistent with a

helical secondary structure (Figure 12 C). The melting temperature as measured by circular dichroism thermal unfolding experiments increases by 5.3 °C when some additional ionic strength is present in the buffer (Figure 12 D).

The detailed biophysical characterization indicates that the NTA is pure, stable and well folded, and thus, thus a suitable sample for crystallization. However, all crystallization attempts failed. Taken the benefit of the invitation to attend an EMBO workshop on "High throughput Protein production and Crystallization" at Harwell (UK), the NTA was used as a test sample in the hope that new high-throughput techniques would lead us to a crystal. During the course several modification to the expression and purification protocol were introduced. This included autoinduction expression, which increased protein yield 4-fold, and new ultrafiltration devices that allowed us to concentrate the protein up to 90 mg/ml. Despite all the variations tried, our attempts did not yield any crystallization hits.

9. NMR structure of MuB N-terminal appendage

As the NTA is quite soluble and of an appropriate size we determined its three dimensional (3D) structure by NMR spectroscopy. The NMR samples used for sequential assignment and structure determination comprise both the MuB construct harboring residues 1-63 expressed in *E.coli* and grown in LB (non isotopic labeled sample) or M9 minimal media using ^{15}N-NH_4Cl as a sole nitrogen source (^{15}N isotopic labeled sample). The assignment of ^{1}H, ^{15}N, and ^{13}C (natural abundance) resonances was accomplished using 500 µM samples in 20mM sodium phosphate buffer pH 7.5 and 100 mM NaCl and 5.5% D_2O with a combination of 2D ^{1}H-^{1}H TOCSY, 3D HNHA, 2D ^{1}H-^{1}H NOESY and 3D ^{15}N-NOESY-HSQC (Figure 13 A). The solution NMR The solution NMR structure was calculated by simulated annealing in torsion angle space using as restraints both upper distances from integrated 2D and 3D NOESY cross-peaks and dihedral ϕ and ψ torsion angles predicted with TALOS+. The calculated structures are very well defined and in excellent agreement with the NMR data (Table 3). No violations greater than 0.20 Å or 5° for the experimental

distances and angle restraints, respectively, were found. Favorable conformational energies and Ramachandran plot statistics (Figure 13 B) resulted for the ensemble of structures.

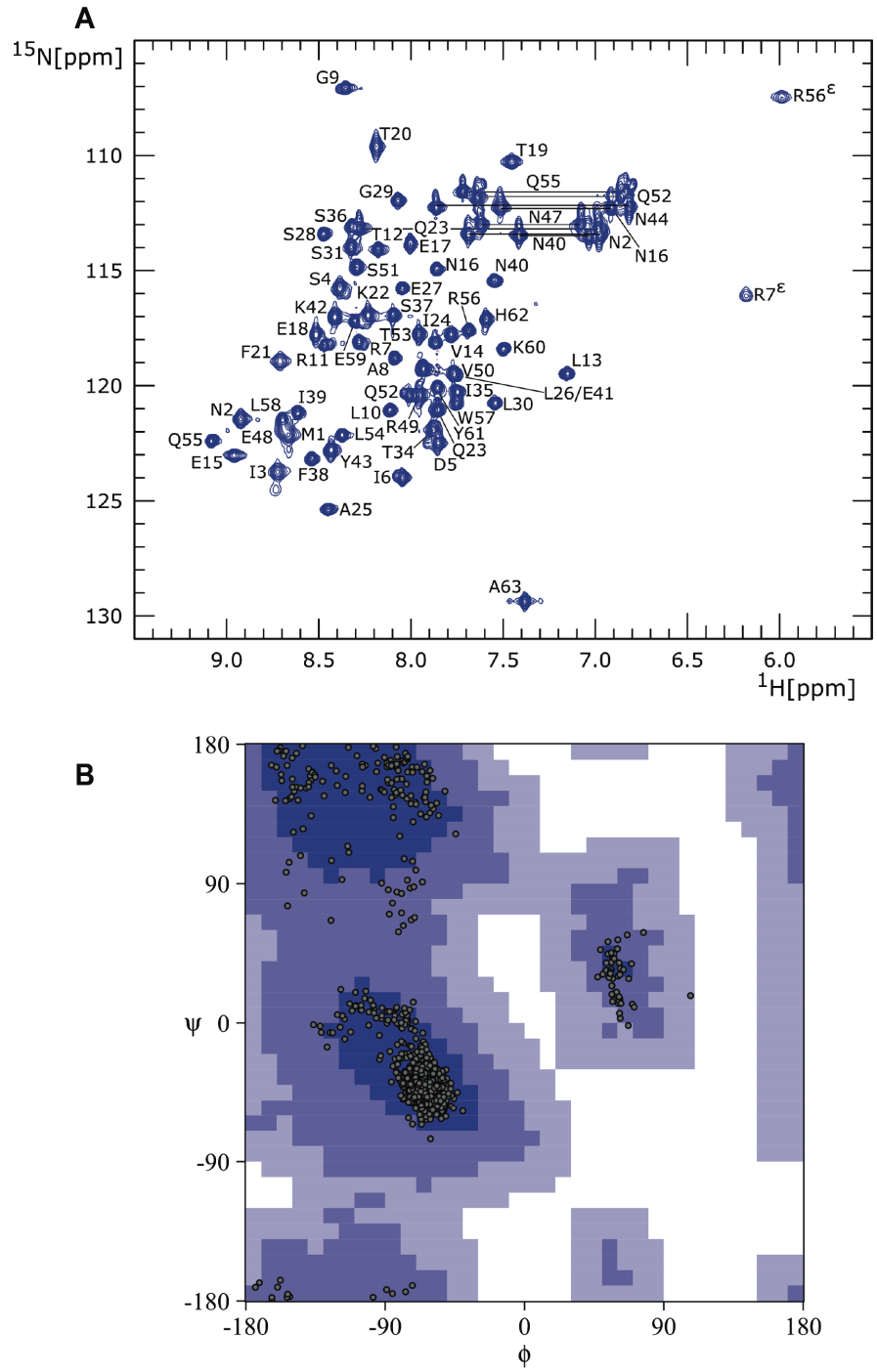


Figure 13. NMR structure of MuB N-terminal appendage. (A) ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N labeled MuB N-terminal appendage (residues 1-63) in 20 mM sodium phosphate pH 7.5 and 100 mM NaCl at 25 °C. Residue assignments are labeled according to sequence numbering for the intact protein. (B) Ramachandran plot for MuB N-terminal appendage.

Table 3. Summary of the experimental restraints and structural statistics for the NMR solution structure of MuB N-terminal appendage.

NOESY cross-peaks [a]		
total	3464	
assigned	3461	99.91%
2D NOESY	2645	99.88%
3D ^{15}N-NOESY-HSQC	816	100%
unassigned	3	0.09%
Distances restraints		
total	1518	
short-range l*i*-*j*l ≤ 1	696	45.85%
medium-range 1 < l*i*-*j*l < 5	390	25.69%
long-range l*i*-*j*l ≥ 5	432	28.46%
Angle restraints (TALOS$^+$)		
ϕ/ψ	46/46	
Structure statistics [b]		
CYANA target function (Å^2)	0.71 ± 0.17	
AMBER energy (kcal/mol)	-2879.94 ± 76.00	
Average RMS deviations from restraints		
distance restraints (Å)	0.0105 ± 0.0004	
angle restraints (°)	0.1691 ± 0.0591	
Average number of violations		
distances restraints > 0.20 Å	0 (maximal 0.16 Å)	
dihedral restraints > 5°	0 (maximal 2°)	
Average RMS deviations from ideal covalent geometry		
bonds (Å)	0.0141 ± 0.0001	
angles (°)	1.663 ± 0.045	
Ramachandran plot [c] (%)	93.1/6.6/0.3/0.0	
RMSD to mean coordinates [d] (Å)	0.41/0.86	

[a] NOESY cross-peaks used to generate distance restraints within the automated NOE assignment algorithm in CYANA.

[b] Calculated over the final bundle of 20 representative conformers after energy minimization with OPALp against the AMBER force field.

[c] Percentage of residues in most favored, additionally allowed, generously allowed, and disallowed regions of the Ramachandran plot.

[d] RMSD values for the backbone atoms N, C^{α}, C′ or for all the heavy atoms, respectively, in the structured region of the protein spanning residues 2-63. To obtain the RMSD value of a structure represented by a bundle of conformers, all conformers are superimposed on the first one and the average of the RMSD values between the individual conformers and their average coordinates are computed.

The precision of the NMR structure, measured by the RMSD value relative to the average coordinates for the backbone atoms N, Cα, and C' from residue N2 to A63 is 0.41 Å and 0.86 Å when all heavy atoms were used. The corresponding values for the backbone and heavy atom RMSD over the four α-helices are 0.29 Å and 0.78 Å, respectively.

Figure 14 A shows the ensemble of the representative set of 20 NMR structures superimposed on the heavy backbone atoms from residue 2 to 63.

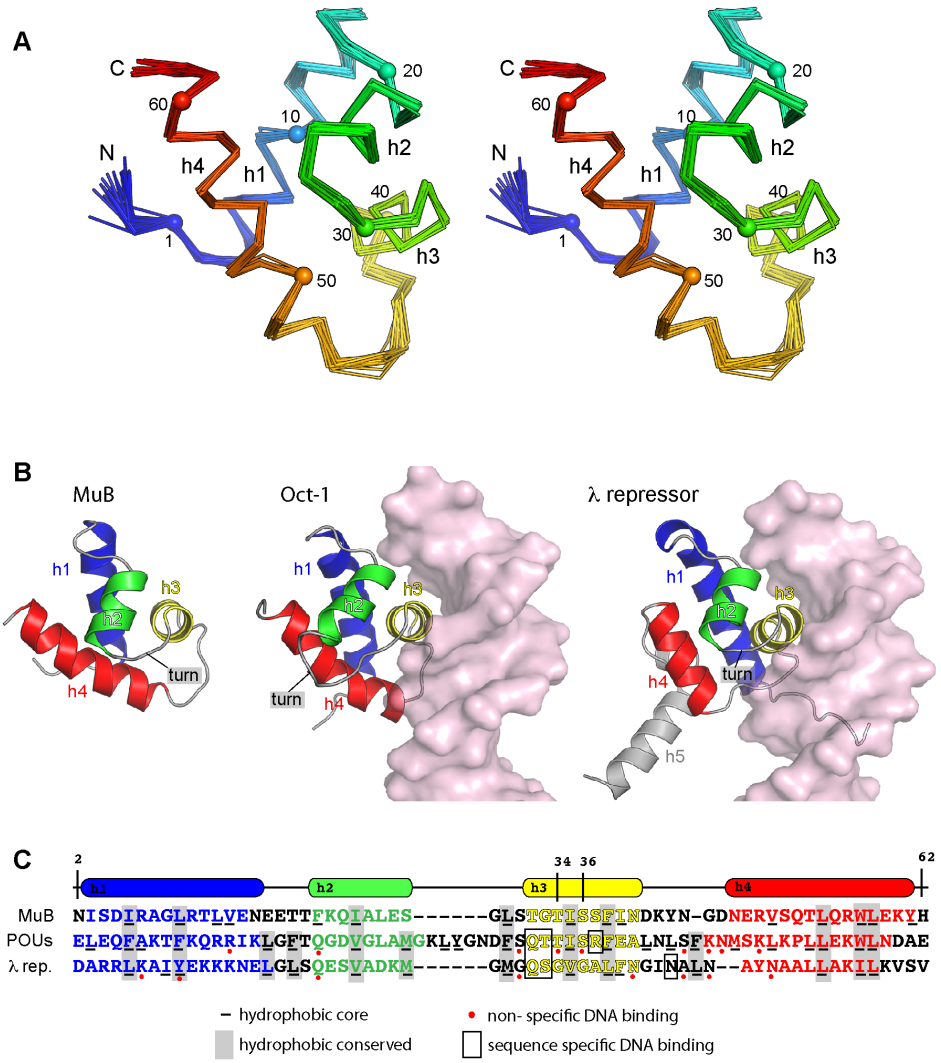


Figure 14. NMR structure of MuB N-terminal appendage and its comparison with λ-repressor like DNA-binding domains. (A) Stereo diagram of 20 superimposed NMR structures of MuB N-terminal appendage. (B) Structural comparison of MuB N-terminal appendage with POUs domain of Oct-1 (PDB 1E3O) and N-terminal domain of λ-repressor (PDB 1LMB) bound to DNA. (C) Structural-based alignment of MuB N-terminal appendage with POUs domain of Oct-1 and N-terminal domain of λ-repressor.

The NTA consists of four α helices. Helices h1 (residues 3 to 15) and h2 (21-28) are the longest, whereas h3 (32-39) and h4 (48-61) are shorter and form a helix-turn-helix (HTH) motif. HTH motifs are commonly found in many DNA binding proteins that regulate gene expression. The motif consists of two α-helices joined by a short turn; the first helix stabilizes the structure whereas the second helix is named "DNA recognition helix" because if recognizes the DNA by fitting into the major groove (Sauer et al 1982). Helices h1 and h4 are aligned in parallel and cross each other forming an angle of 50°. Helix h2 forms a kink with h1 and packs against h4 and is a nearly perpendicular orientation. There is a short linker between helices h2 and h3 that makes a sharp turn forming the characteristic HTH motif. The final helix h4 is connected to h3 by a long linker, which extensively interacts with the other three helices. The four helices in the NTA pack tightly enclosing a hydrophobic core comprised by residues I6, L10, L13 and V14 from helix h1; F21 and I24 from helix h2; L30 from the loop h2-h3; I35, F38 and I39 from helix h3; and V50, L54, W57, L58 and Y61 from helix h4 (Figure 14 C). The protein surface has a hydrophilic character, with moderately abundance of charged residues. The computed surface potential shows an acidic patch localized at the protein face where the C-terminus of helix h4 exits to the protein surface.

10. MuB N-terminal appendage shows structural similarity to DNA-binding domains

Comparison of the NTA structure with other known structures in the PDB using DALI server (Holm and Sander, 1993) showed a striking resemblance to the λ repressor-like family of DNA binding domains. This family includes HTH-containing proteins that adopt a similar topology and include phage repressors λ C1 and Cro, bacteriophage 434 C1 and Cro, P22 C2, and the POU specific domain (POUs) of human transcription factor Oct-1.

Comparison of the NTA and POUs structures yields an RMSD of 2.5 Å for the superposition of 60 Cα atoms, whereas the superposition with the λ repressor shows an RMSD of 2.6 Å for 53 Cα atoms. The relative orientation of the helices in NTA is

more similar to POUs than to the λ repressor, particularly for the orientation of helices h1 and h4 (Figure 14 B). These helices cross each other both in NTA and POUs, whereas in the λ repressor they are oriented in the same direction and do not cross. Neither MuB NTA nor POUs present the C-terminal fifth helix that is responsible for the dimerization of the λ repressor. This helix is replaced by a linker sequence that in POU connects the POUs and POUh domains, whereas in MuB connects the NTA with the AAA+ module.

The POUs sequences exhibit a high conservation in the residues forming the hydrophobic core (Figure 14 C). These residues are also conserved in MuB and in λ repressor. The sequence conservation in h3, the "DNA recognition helix", between NTA and POUs is also significant. NTA shares 4 out of the 9 residues that are invariant among POUs domains (Assa-Munt, Cell, 93), including two residues (NTA residues T34 and S36) that in POUs are involved in non-specific DNA interactions. The NTA also has two residues outside helix h3 (S31 and R49) occupying the equivalent POUs position of a serine and a lysine that participate in non-specific DNA binding. However, other residues that in POUs and λ repressor are involved in direct readout of nucleotide bases are not preserved in MuB.

The similarities of the overall architecture of NTA with POUs and λ repressor and the conservation of some of the DNA interacting residues strongly suggest a possible role of the NTA in DNA recognition. To test this hypothesis, we measured 1H-15N HSQC spectra of the protein in the presence and absence of double and single stranded short DNA fragments. However, the lack of changes in the chemical shifts indicated that there was not interaction between the NTA and the DNA (data not shown). We also failed to prove the interaction by using radiolabeled DNA in electrophoretic mobility shift assays (data not shown). Further experiments will be needed to test whether or not the NTA is a DNA-binding domain.

Discussion

1. Architecture and domain functions of MuB

The AAA+ ATPase superfamily (PFAM PF00004) groups more than 35,000 proteins distributed in all kingdoms of life and involved in many cellular such as DNA replication, regulation of gene expression, protein proteolysis and disaggregation, intracellular transport, microtubule severing, peroxisome biogenesis, membrane fusion or signal transduction (Erzberger and Berger, 2006). These proteins are often referred as molecular ''machines'' for the efficiency with which they convert the chemical energy provided by ATP hydrolysis into mechanical manipulation of a great diversity of protein or DNA targets. The defining characteristic of AAA+ is a conserved module of ~230 residues that binds and hydrolyzes ATP (Figure 15).

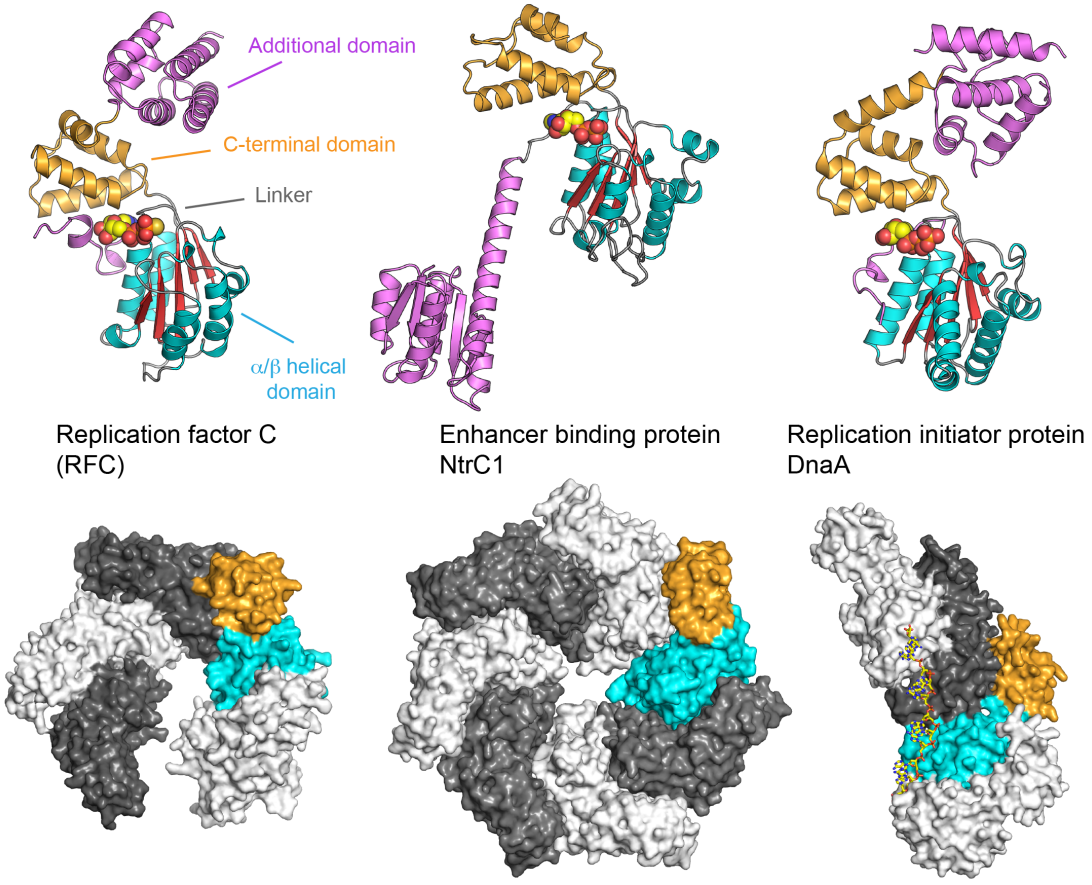


Figure 15. AAA+ ATPases. Reprezentation of the AAA+ fold exemplified by the oligomers and subunit structures of: Replication factor C RFC (PDB 1SXJ), Enhancer binding protein NtrC1 (PDB 1NY5) and Replication initiator protein DnaA (PDB 3R8F). The α/β domain is represented in cyan and red, C-terminal domain in orange and an additional domains in magenta.

The AAA+ module contains two domains: an N-terminal domain composed of a 5-stranded parallel β-sheet flanked on both sides by α-helices, and a C-terminal helical bundle. The absence of additional β-strands adjacent to the 5-stranded β-sheet and the presence of a helical bundle attached to the C-terminus of the α/β domain are primary features that distinguish AAA+ proteins from other P-loop nucleotide-binding proteins. Some AAA+ have additional regulatory domains added at the N-terminus (as in MuB) or at another site, and can be integrated in larger molecular assemblies (Figure 15), which further increases their structural and functional diversity (Hanson and Whiteheart, 2005).

Typically, AAA+ proteins bind ATP in a cleft formed between the carboxy-terminal edge of the central β-sheet and the closure of the C-terminal helical domain (Figures 7 C and 15). Each domain provides characteristic elements for nucleotide binding and hydrolysis, and their spatial positioning is highly conserved. Some of these elements, the Walker A (or P-loop), located between strand β1 and helix α1, and the Walker B, at the carboxy-end of β3, are conserved with other nucleotide-binding proteins. Contrary, the nucleotide-interacting elements sensor-I and sensor-II are exclusive of the AAA+ ATPases. In addition, the AAA+ active site is often completed by at least one arginine residue from an adjacent subunit –the R finger– that interacts with the ATP. Hence, AAA+ proteins are active as oligomers, and they typically assemble into ring-like or helical oligomers with a central hole to accommodate the substrate (Figure 15).

In this work, we combined sequence alignment, site-directed mutagenesis, and biochemical assays to prove that MuB folds into a *bona fide* AAA+ module, with an N-terminal α/β domain (residues 77-213) and a C-terminal helical domain (residues 231-312) connected by a linker that matches the sequence prone to proteolytic cleavage. The discovery of ATP-induced MuB helical filaments visualized by EM provides further support to the assignment of MuB as an AAA+ ATPase. We also demonstrate that the N-terminal ~70 residues of MuB constitute a topologically independent domain, loosely attached to the AAA+ module, which appears to play a relatively minor role in the known biochemical activities of MuB and is not required for filament formation. This new picture of MuB architecture explains the difficulties previously encountered to characterize its function. Studies with full-length MuB

were hampered by the poor solubility of the protein in presence of ATP (Chaconas et al., 1985; Teplow et al., 1988) a behavior that is now understood by observing the rapid assembly of MuB into filaments and the tendency of these filaments to aggregate into large bundles. On the other hand, it also explains why previous attempts to studying the proteolytic fragments of MuB provided little understanding of the protein function and failed to trace the functional elements. The two AAA+ domains act as a single unit and we demonstrate that a complete MuB AAA+ module is needed and sufficient for ATP binding/hydrolysis and for filament assembly.

2. Identification of key residues for MuB filament formation

The redefinition of MuB modular architecture guided us in the identification of the characteristic AAA+ functional elements: sensor I and II, and R finger. Mutation of these residues strongly affects ATP binding/hydrolysis. Furthermore, those mutants that cannot bind ATP also loose the capacity to form filaments, proving that ATP glues the subunits together and triggers filament formation.

Sensor-I is typically an asparagine or another polar residue (Ser, Thr, Asp) sitting at the apex of strand β4 in an intercalated position between the Walker A and Walker B motifs. Sensor-I is proposed to coordinate together with the Walker B motif the binding of the Mg^{2+} ion and to properly orient a water molecule for nucleophilic attack on the γ-phosphate of ATP. From the sequence alignment, the role of MuB sensor-I was undoubtedly assigned to N202. As predicted, the mutation N202A strongly impairs ATP hydrolysis, without affecting ATP binding or filament formation.

Sensor-II is typically an arginine whose side chain extends from the C-terminal domain toward the β- and γ-phosphates of ATP. Thus, mutations in sensor-II have a strong effect on ATP binding and hydrolysis. We proposed that R268 is the sensor II of MuB. Indeed, mutation R268A impairs ATP binding and hydrolysis activity, and no filament formation takes place. The fact that this mutation has the lowest ATPase activity of all the mutants studied here is intriguing, since mutating sensor II in other AAA+ proteins causes only a modest decrease in ATPase activity or has no effect in oligomerization (Wendler et al., 2012). This suggests that the C-terminal domain of

MuB might be flexible and that R268 plays a key role in coupling the ATP binding to locking the C-terminal domain in a conformation that favors protein-protein interactions and filament assembly.

The R finger projects from a conserved position on helix α4 from one subunit and interacts with the γ-phosphate of the ATP bound to the adjacent subunit. Thus, ATP glues neighboring subunits together, favoring and in some cases, as in MuB, triggering polymerization. In most of AAA+ proteins, mutation of the R-finger usually impairs ATP hydrolysis and can affect oligomerization. We propose that R224 plays the role of R-finger in MuB, and indeed, the mutation R224A severely impairs ATP binding and filament formation. However, AAA+ members typically have other arginine residues located on the same surface as the R-finger that also point towards the active site in the neighboring subunit. We mutated two additional arginines in MuB that, according to the 3D model built by PHYRE, were predicted to be on the same protein surface as the R-finger. R220 occupies an equivalent position to an extra arginine in helix α4 of DnaA that projects parallel to the R-finger and forms a salt bridge with sensor-I in the adjacent subunit. The mutation R220A in MuB impairs ATP binding and filament formation. On the other hand, R187 aligns with an arginine that in RFC directly interacts with the γ-phosphate of ATP. In agreement with this, MuB R187A can bind but not hydrolyze ATP. The R187A mutant is also compromised in filament formation, which in this case cannot be assigned to a defect in nucleotide binding. Thus, we propose that an electrostatic interaction between R187 and the ATP γ-phosphate may contribute to stabilizing this protein interface. Oligomerization of other AAA+ members has been shown to be sensitive to salt concentration, indicating the importance of electrostatic interactions. Likewise, MuB filament formation is salt concentration-sensitive. We propose that these three arginines cooperate in stabilizing the polymerization interface in the presence of ATP. Single mutant proteins, although compromised, can still form filaments on DNA, which can interact with MuA and activate the target DNA for strand transfer in the presence of ATPγS, albeit very weakly.

3. Identification of key residues for DNA and MuA binding

The sequence alignment with NtrC1, PsPF and ZraR led us to the identification of the DNA-binding motif of MuB. These enhancer-binding proteins have a conserved sequence in the loop (loop-1) bridging the two halves of helix α2 that is important for interaction with the σ-subunit of RNA polymerase (Rappas et al., 2005). When these proteins assemble into heptameric rings (Figure 15, middle panels), loop-1 from each subunit points to the central pore where the polymerase binds. In the sequence alignment, we noticed that MuB also exhibits a loop that could be inserted within the two halves of the predicted helix α2. The sequence of this loop ^{140}PRRKGP^{154} contains a stretch of positively charged residues and by similarity with the enhancer-binding proteins, we hypothesized that it could point towards the central channel of the MuB filaments and interact with the DNA. Indeed, the molecular docking of NtrC1 AAA+ module within the EM envelop of the MuB filament showed that loop-1 of NtrC1 fitted nicely in the density protruding towards the axial channel (Box 5). The results obtained with the double and triple loop-1 mutants support this idea because their ATPase activity is not inhibited by the addition of DNA and the filaments fail to assemble on the DNA. Synergistic effects of two mutations, R151A and K152A, are needed to eliminate DNA recognition because point mutations do not abolish DNA binding.

Interestingly, mutations in the DNA-binding loop-1 unexpectedly increase the ATPase rate by nearly 2-fold. In addition, it was well known that the addition of DNA lowers the apparent K_d for ATP and slows down ATP hydrolysis (Maxwell et al., 1987; Adzuma and Mizuuchi, 1991). These two results together strongly suggest that DNA binding exerts an allosteric control on the ATPase active site favoring tighter ATP binding and slowing down hydrolysis. According to the 3D model of MuB, the DNA-binding loop and the ATP active site are ~30 Å apart. Without more detailed information about MuB structure it is not possible to envision how the communication between the two sites can take place. Nevertheless, the structures of NtrC1, PsPF and ZraR free or bound to different nucleotides have revealed a nucleotide-dependent movement of loop-1 that might control the interaction with

the σ-subunit (Rappas et al., 2006). A similar mechanism in MuB would explain why mutations in loop-1 could uncouple DNA recognition and the ATPase cycle, allowing for a faster hydrolysis. The stabilization of the ATP-bound state with hydrolysis inhibition promoted by the binding to the DNA would place the DNA-bound MuB ATPase under stringent control of MuA.

MuB forms a tight structure that wraps around the DNA, and therefore, MuA must contact the MuB-DNA filament from the outside. Our results strongly suggest that this interaction involves the three sequential lysine residues at the linker connecting the α/β domain with the C-terminal helical bundle. The multiple sequence alignment shows that other AAA+ members (e.g. NtrC1) also present a linker with positively charged residues (Figure 7 B). At least for some of these proteins, the crystal structures determined both in absence and presence of nucleotides have shown that the linker with positively charged residues undergoes conformational changes associated with the ATPase cycle (Rappas et al, 2006). Thus, an attractive possibility is that the ATPase cycle of MuB could be controlled by conformational changes at this linker region. The fact that the triple linker mutant K233A/K235A/K236A exhibits higher ATPase activity, as the loop-1 mutants, and fails to respond to the activation by MuA strongly support that the exposed linker connecting the two domains of MuB AAA+ module are involved in the interaction with the transposase.

4. MuB filaments: some (dis)assembly required

The EM and biochemical data of MuB and the characterization of a battery of MuB mutants together with the accumulated knowledge on AAA+ ATPases help us to put the pieces together on how ATP induces the MuB filaments assemble on the DNA and how MuA promotes its disassembly (Figure 16).

MuB is a monomeric protein with a dynamic N-terminal appendage and an AAA+ module. The linker connecting the N- and C-domains of the AAA+ module is probably exposed and highly flexible, which would explain its extreme susceptibility to proteolytic cleavage. Thus, in the absence of ATP, MuB is a very flexible protein, and perhaps a hopeless target for crystallization. In this flexible state, MuB can probably

bind to the DNA through loop-1, but the interaction is likely to be weak. The binding of ATP between the α/β domain and the C-terminal bundle fixes the AAA+ module in a conformation that favors MuB oligomerization. Perhaps the formation of small MuB oligomers with an increased affinity for the DNA –provided by the sum of the small affinities of each subunit– precedes binding to the DNA, in agreement with previous fluorescent microscopy observations that suggested the recruitment of a protomer of MuB to the DNA (Greene and Mizuuchi, 2002a). We hypothesize that the size of the immediate precursor for DNA binding must be not larger than 5 MuB subunits, otherwise the filament would form a complete helical turn and the DNA could not enter easily in the axial channel. Upon DNA binding, loop-1 might adopt a conformation that decreases the ATPase rate, and thus, favors the stability of the filament. Then, the filament extends by the coupling of additional MuB-ATP subunits.

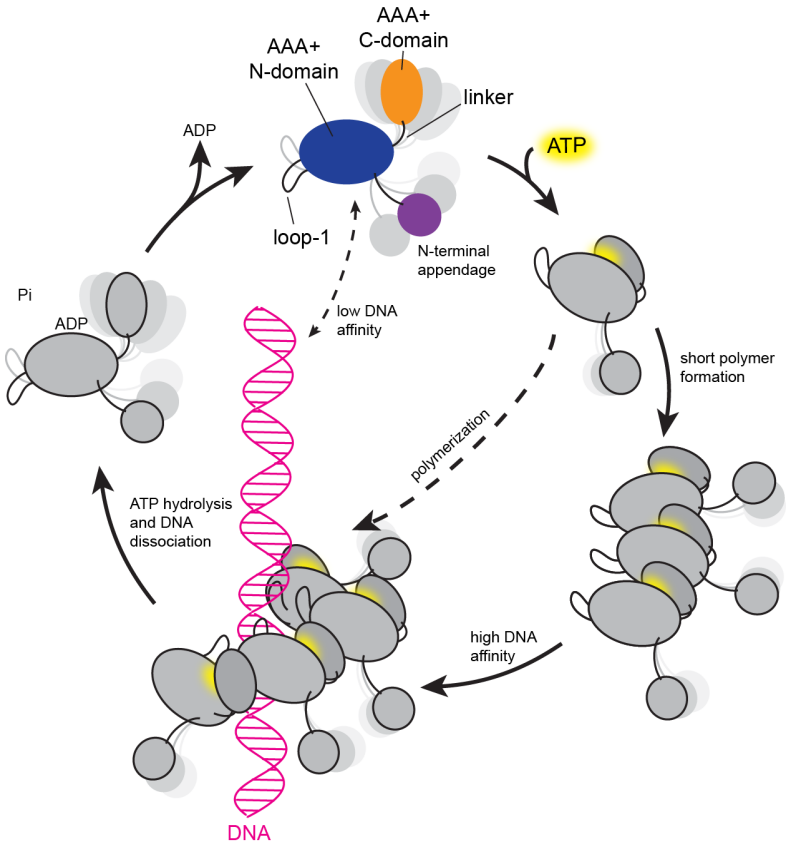


Figure 16. Mechanistic model of the assembly and disassembly of MuB filament on the DNA.

According to the EM reconstructions, each MuB subunit interacts with its immediately adjacent subunits, involving the R finger and residues R187 and R220, but not with subunits in the adjacent spirals of the helix. We believe that the lack of interaction between spirals along the filament axis might be the reason behind the intrinsic flexibility of the filaments observed by EM. These small variations within the number of subunit per turn made the 3D reconstructions of the filaments quite challenging. If no classification of the filament particles was done, the 3D reconstructions showed a smooth and featureless spiral, and only by exhaustively classifying the images the 3D reconstructions for some of these classes showed well-defined subunits. Nevertheless, even within the small variations observed for the protein filaments, we can confirm that the parameters of the MuB helix differ from those of the DNA within the filament. This implies that the interactions of loop-1 with the DNA change along the filament axis with the consequent variations on the ATPase activities. This is perhaps an important contributing factor for the formation of short rather than long MuB segments. Fluorescent experiments indicated that the size of the polymers varies between 10-60 MuB subunits (Greene and Mizuuchi, 2002b), which would correspond to ~2-11 helical turns. The relatively weak cooperativity in MuB-DNA polymerization together with the weak DNA binding might be advantageous so that MuB can be easily dispersed by MuA coupled ATP hydrolysis during transposition immunity, and perhaps it also favors the displacement of MuB filaments by other DNA interacting proteins, so that MuB does not interfere too much with normal cellular processes.

The dissociation of MuB filaments from the DNA is prompted by the interaction with MuA with the linker region. We propose that conformational changes in this loop might stimulate ATPase activity. ATP hydrolysis would trigger the rapid dissociation of MuB from the DNA due to the change of the DNA binding loop to a low affinity conformation, the weakened contacts between the R finger and the released γ-phosphate of ATP, and the recovered flexibility between the two halves of the AAA+ module. The released MuB subunit will be free to exchange the ADP by ATP and start a new cycle of polymerization.

5. How does MuB target DNA for transposition?

The mechanistic description of MuB-DNA assembly and disassembly is a nice model that synthesizes a lot of the acquired biochemical knowledge about MuB. However, it does not explain how MuB converts the DNA into a better substrate for transposition. Under the light of this model, we re-analyzed unpublished data obtained from the group of Kiyoshi Mizuuchi, where they investigated the size of MuB-DNA filament needed to stimulate MuA strand transfer reaction (Box 1). Although at that time, it was not know that MuB binds the DNA by forming a protein crust around it, the results clearly showed that the efficiency of the transposition increases with the concentration of MuB, but that when the DNA is fully covered by MuB, it becomes a poor substrate for transposition. Hence, we conclude that the DNA within the MuB filament is not accessible to MuA, and that strand-transfer reaction must be favored at DNA sites adjacent to the filament ends.

To understand how MuB filament ends can present the DNA to the transposase, it is important to emphasize an important feature of MuB-DNA filaments: MuB helix wraps the DNA without seemingly altering its structure, and thus, the helical parameteres of the MuB polymer do not match those of the B-form DNA (Figure 17 A). This symmetry mismatch is a feature that, to our knowledge, has not been previously observed in other nucleoprotein filaments. Commonly, protein filaments that contain DNA either adopt a helical symmetry close to that of the standard B-form DNA, or impose their symmetry on the DNA, inducing deformations. Well-known examples are the RecA family of ATPases, which assemble into helical filaments on DNA and catalyze homologous DNA pairing and strand exchange. Each monomer binds three nucleotides, while extending and untwisting the DNA between the triplets, causing an overall stretching of the DNA molecule (Chen et al, 2008) (Figure 17 B). Another example is the bacterial DnaA protein, an AAA+ member that assembles into right-handed helices and induces the deformation of DNA in a way similar to the RecA mechanism (Duderstadt et al., 2011) (Figure 17 C).

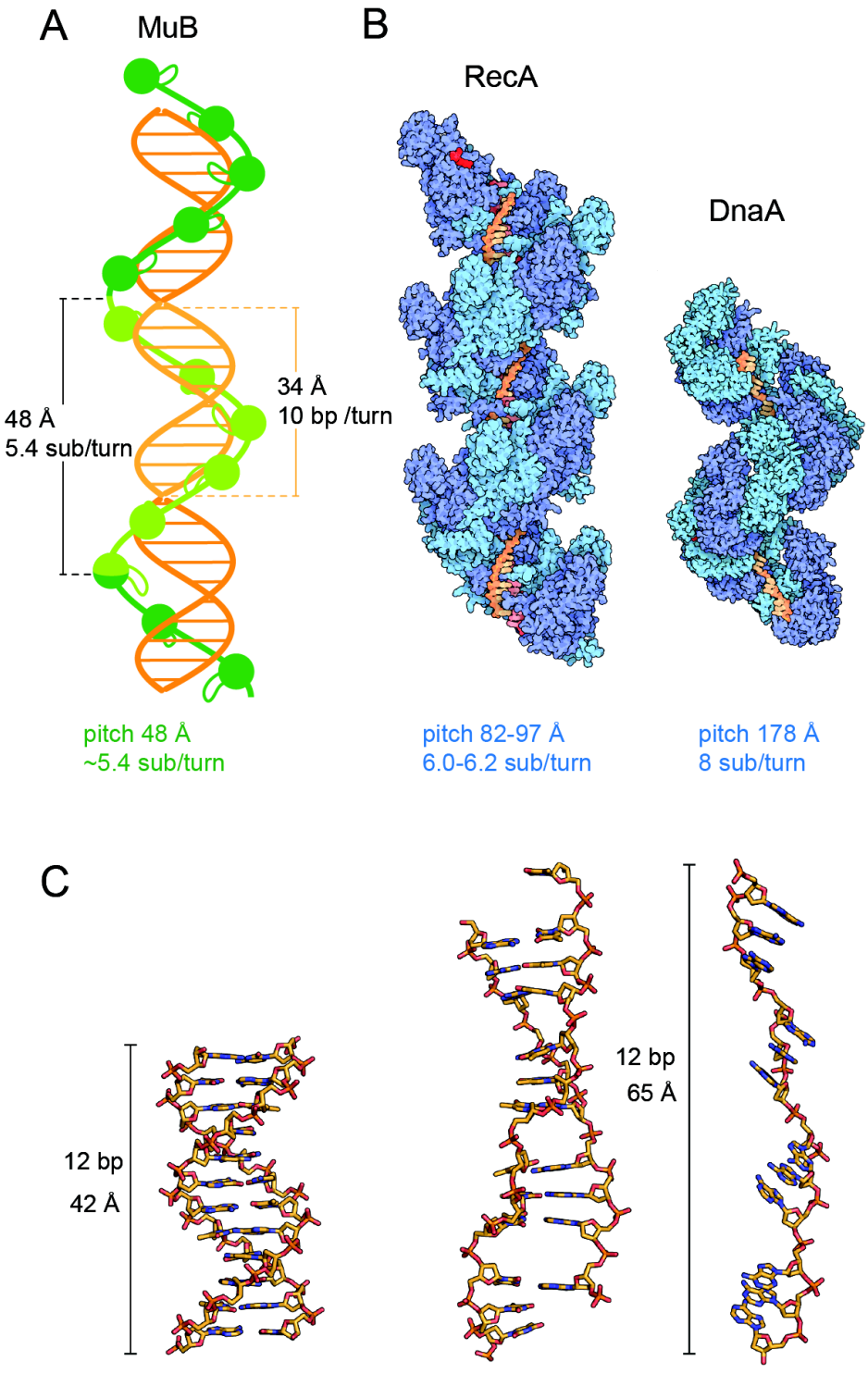


Figure 17. Symmetry mismatch in MuB-DNA nucleoprotein filament. (A) The helical parameters of the MuB filament do not match those of B-form DNA. (B) RecA and DnaA bound to single and double stranded DNA and their helical parameters. (C) Comparison of the DNA bound to MuB, DnaA and RecA.

However, not all nucleoprotein filaments induce DNA deformation. RFC is a complex of five different AAA+ subunits that forms right-handed spirals with roughly the same pitch as B-form DNA, without causing apparent deformations (Bowman et al, 2004). An interesting example is ParA2, a bacterial protein involved in plasmid segregation, which forms left-handed helical filaments on DNA without significantly changing the B-form DNA helicity (Hui et al., 2010). Unlike RFC, ParA2 filaments do not match the DNA helical parameters, but the arrangement of the helix is dictated by the repetitive interaction of the protein every 8 bp in the DNA, resulting in two different but complementary helices.

The lack of symmetry between the MuB and DNA helices implies that the protein-DNA contacts vary from point to point along the filament, as mentioned before. A possible consequence of this irregular association with DNA is that not all MuB subunits have loop-1 in the same conformation, and thus the ATPase rate might vary among them. One might ask if the stimulation of MuB ATPase by MuA could be transmitted to the DNA-binding loop, changing its conformation and inducing local changes of the DNA structure. In fact, the most efficient stimulation of MuB ATPase requires a tetramer of MuA molecules bound to the Mu DNA ends (Mizzuchi, 1992; Lavoie and Chaconas, 1995), which suggests that multiple nearby MuB monomers might be activated simultaneously. Could this impose local symmetry-matching and DNA deformation that in turn might help utilization of this DNA segment as the strand transfer substrate for MuA? Although MuA lacks sequence specificity, it is known to display a great preference for insertion in mismatch sites on the DNA (Yanagihara and Mizuuchi, 2002). There is no physiological reason for such preference other than that the mismatch DNA would readily assume a deformed DNA structure, which is favored for the strand transfer reaction. Indeed, the target DNA segment is severely bent within the structure of the Mu strand transfer complex (Montaño et al, 2012). Since MuB forms a tightly packed outer shell, the DNA inside the filament would be inaccessible for MuA strand transfer. Indeed, MuB-saturated DNA is a poor target for Mu transposition. Thus, probably, the DNA selected for Mu insertion is not encased within the filament, but it is at the MuB filament boundaries. Then, the hypothetical DNA deformation discussed above

perhaps takes place when MuB disassembly is triggered by the MuA-Mu end complex, ready for strand transfer (Figure 18).

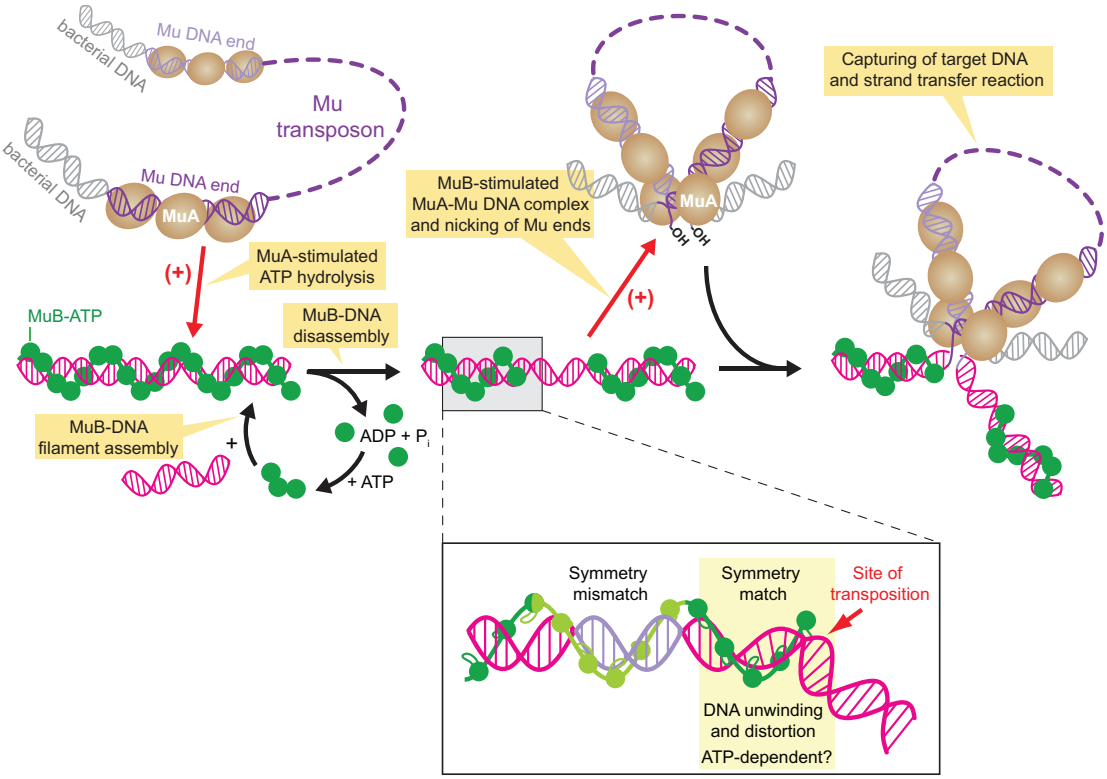


Figure 18. Model for MuB targeting DNA for transposition. Upon ATP binding, MuB forms helical filaments on DNA. MuA bound to Mu DNA ends stimulates ATP hydrolysis by MuB and MuB dissociation from DNA, which generates MuB-free DNA regions. Reciprocally, MuB stimulates MuA to pair and nick Mu DNA ends at the junction with the flanking sequences. MuA and MuB together may induce the matching symmetry between MuB and DNA at the boundary of a MuB filament and thus DNA distortion, which leads to the target DNA capture and Mu transposition.

Short helical MuB filaments might be too unstable on DNA and need a minimum of several helical turns to avoid premature dissociation for productive interaction with the MuA-Mu end complex for strand transfer activation. Perhaps there is an optimal polymer size distribution for function, and the MuB architecture described here is well tuned toward this, and also to respond efficiently to its interaction with MuA at different stages of the transposition reaction.

6. N-terminal appendage-mediated filament bundles might confer target immunity

As it was explained in the introduction, there are two mechanisms to explain Mu target immunity that appear to be incompatible between them. The target mechanism assumes that MuB is depleted by MuA from the regions near the initial insertion site and accumulates at distant DNA regions that are presented by MuB as better substrates for transposition. On the other hand, the alternative mechanisms proposed by Harshey (see introduction) the opposite, that MuB is highly enriched in the Mu genome and that together with some bacterial proteins segregates the Mu genome as a different territory that is not recognized by the transposase. Which of the two models is the correct one? Although apparently one model excludes the other could they both be correct?

Altogether our results support that MuB can have a double role, playing offense and defense on DNA transposition. Could this double personality of MuB reconcile the two proposed mechanism of Mu phage target immunity? The active role of MuB in targeting DNA for the transposase has been widely studied, and our results propose a mechanism by which MuB filaments promote DNA deformation and delivery to MuA for transposition. However, we also demonstrate that the architecture of MuB around the DNA makes it of difficult access to the transposase and the DNA fully coated by MuB is a poor substrate for transposition. Could this protection account for the new mechanism of immunity? The protection of the DNA by the formation of a linear MuB filament, as observed by EM, would be ineffective, since the amount of MuB required to cover the multiple copies of the 37 kb Mu genome would likely exceed those present during viral infection. Therefore, the images of DNA saturated with MuB are not likely to happen in vivo, and a more realistic scenario would be that shown in figure 7 D with patches of filaments distributed at different positions along the DNA. Nevertheless, could there be a mechanism by which a limited number of scattered MuB segments could occlude lengthy DNA sequences from the action of the transposase? One possibility is that the observed tendency of the filaments to stick to each other and form bundles is a physiologically relevant

characteristic of MuB. In this way, the total amount of MuB available is not critical, since a limited number of more or less short filaments distributed along the DNA could interact with each other, compacting the DNA into a condensed state that might turn into a poor substrate for transposition (Figure 19).

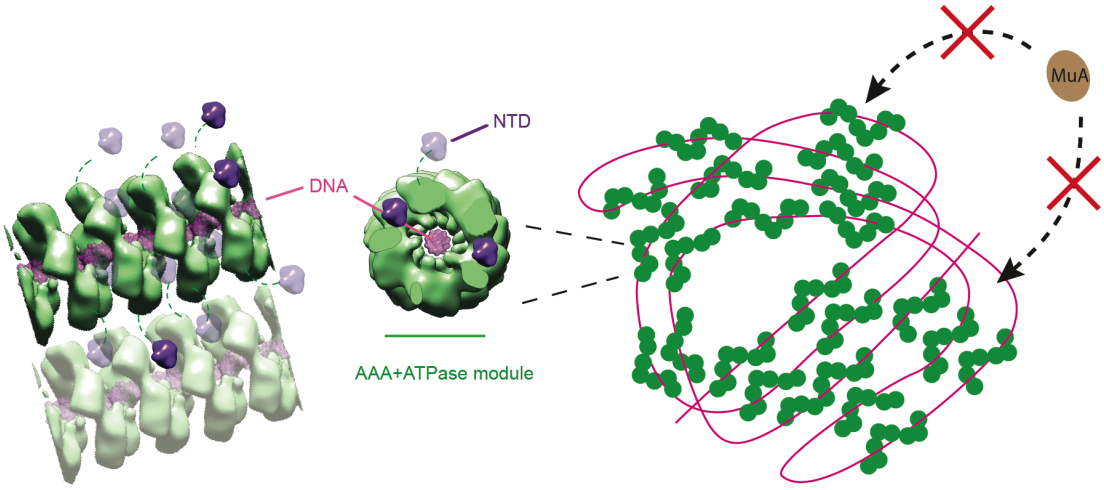


Figure 19. Model for MuB target immunity. MuB filament-filament interactions trough N-terminal appendage promote DNA condensation and provide protection against the action of MuA transposase.

A similar strategy is used during retroviral infection, where the cellular protein BAF (Barrier to autointegration factor) binds and occludes the viral DNA protecting it from self-integration (Zheng et al, 2000; Bradley et al., 2005). Thus the poor solubility of the MuB-ATP assemblies, which gave so many troubles for the characterization of MuB function, could indeed hide an important cross-bridging capability that might protect the Mu genome from the transposase. Filament stickiness had not been observed by fluorescent microscopy. One could argue that the EGFP attached to the N-terminus of MuB must provide the filament with an outer crust that would prevent the interaction with other filaments. Indeed, we have observed that EGFP-MuB forms filaments in the presence of ATP, but these filaments do not aggregate into bundles (data not shown). Interestingly, we did not observe this filament stickiness neither with MuB-ΔN, suggesting that perhaps the N-terminal 70 residues appended to the AAA+ module could be involved in filament-filament interactions.

To better understand the role of the N-terminal appendage we determined its NMR structure. The NTA is a compact domain, sufficiently small to fit in between the spirals of the MuB filament. Perhaps the NTA projecting from one filament could move inside an adjacent filament. The linker connecting the NTA and the AAA+ fold is 14 residues long and predicted to be disordered, and thus, if fully extended it could reach a neighboring filament. One of the 3D reconstructions of MuBwt without DNA shows a density that occupies the axial channel (Box 3). One possibility is that this corresponds to the NTA, which in the absence of the DNA bind occupies partially the axial channel. This would support that the NTA fits within the spirals of the helix and could reach the position of the DNA (Figure 19). The striking structural similarity between NTA and POUs, and the conservation of most of the residues involved in non-specific interactions support this notion. The fact that we were unable to detect the interaction of NTA with DNA using NMR or EMSA approaches does not discard this hypothesis. In fact, the POUs domain shows very week affinity for the DNA when isolated from the POU homeodomain. We would expect that because NTA lacks some of the residues that in POUs make specific interaction, the DNA binding would be even weaker and more difficult to detect. However, it is possible that the strong tendency of the MuB filaments to aggregate could be promoted by the sum of numerous NTAs cross-interacting between filaments, similarly to a Velcro-like mechanism.

For many years, it has been known that RecA formed bundles with a tendency to aggregate into bundles (Brenner et al., 1988). This bundles, however, were thought to be artifacts or, in the best of the cases, some short of reservoirs to store the proteins in an inactive state. However, recent reports have reported important roles of these bundles *in vivo* (Lesterlin et al., 2014). We have also spent a long time believing that the bundles were an artifact to be avoided. If and how the NTA induced MuB filament bundles occur in vivo, and what would be the importance of the bundles for occluding the DNA from the action of the transposase are important questions that remain to be answered.

Conclusions

Conclusions

1. MuB is a previously unrecognized member of the AAA+ ATPase superfamily.
2. MuB is composed of two domains: (1) an N-terminal appendage and (2) an AAA+ motif divided into an α/β domain and a C-terminal helical bundle.
3. Upon ATP binding, MuB forms helical filaments that wrap around single or double stranded DNA.
4. We identify the MuB residues involved in ATP binding and hydrolysis, filament formation, DNA binding and interaction with MuA.
5. We propose a mechanistic model of how MuB forms filaments on the DNA in an ATP-dependent manner and how the interaction with MuA triggers the dissociation of the filaments.
6. MuB filament helical parameters differ from those of B-form DNA, but the formation of the nucleoprotein filament does not seemingly alter the structure of the DNA. This unique symmetry mismatch suggests a mechanistic model of how MuB could deform the DNA at the filament end making it a better substrate for transposition.
7. MuB filaments form bundles that depend on the presence of the N-terminal appendage.
8. We determined the structure of the N-terminal appendage by NMR spectroscopy. The structure is strikingly similar to those of the λ-repressor like DNA-binding domains, suggesting that the N-terminal appendage could be involved in DNA binding.
9. We propose a model on how MuB filament-filament interactions trough N-terminal appendage could be involved in an alternative mechanism of target transposition immunity.

Conclusiones

1. MuB es un nuevo miembro de la superfamilia de las ATPasas AAA+.
2. MuB está formado por dos dominios: (1) un apéndice N-terminal y (2) un motivo AAA+ dividido en un dominio α/β y un conjunto de hélices C-terminales.
3. Al unir ATP, MuB forma filamentos helicoidales que pueden enrollarse entorno al ADN de simple y doble cadena.
4. Hemos identificado los residuos de MuB involucrados en la unión e hidrólisis del ATP, en la formación de filamentos, la unión del ADN y la interacción con MuA.
5. Proponemos un modelo mecanístico de cómo MuB forma filamentos sobre el ADN en una manera dependiente de ATP y cómo la interacción con MuA promueve la disociación rápida de los filamentos.
6. Los parámetros helicoidales del filamento de MuB difieren de los del ADN con forma B, pero la formación del filamento nucleoproteíco no parece alterar la estructura del ADN. Este mal emparejamiento no se ha observado previamente y sugiere un mecanismo de cómo MuB podría deformar el ADN en los extremos del filamento, convirtiéndolo en un mejor sustrato para la transposición.
7. Los filamentos de MuB forman haces que dependen de la presencia del apéndice N-terminal.
8. Hemos determinado la estructura del apéndice N-terminal mediante espectroscopía de NMR. La estructural es sorprendentemente similar a las de los dominios de unión a ADN de la familia de los represores λ, sugiriendo que el apéndice N-terminal podría estar involucrado en la unión al ADN.
9. Proponemos un modelo de cómo las interacciones entre los filamentos de MuB mediadas por los apéndices N-terminales podrían proporcionar un mecanismo alternativo de immunidad frente a la transposición.

References

Adzuma, K. and Mizuuchi, K. (1988). Target immunity of Mu transposition reflects a differential distribution of Mu B protein. Cell 53, 257–266.

Adzuma, K. and Mizuuchi, K. (1991). Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage Mu. Involvement of protein oligomerization in the ATPase cycle. J. Biol. Chem. 266, 6159–6167.

Ammelburg, M., Frickey, T. and Lupas, A. (2006). Classification of AAA+ proteins. J. Struct. Biol. 156, 2–11.

Baker, T.A., Mizuuchi, M. and Mizuuchi, K. (1991). MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell 65, 1003–1013.

Baker, T. A., and Mizuuchi, K. (1992). DNA-promoted assembly of theactive tetramer of the Mu transposese. Genes Dev. 6, 2221-2232.

Baker, T.A., Mizuuchi, M., Savilahti, H. and Mizuuchi, K. (1993). Division of labor among monomers within the Mu transposase tetramer. Cell 74, 723–7.

Baker, T. A., Kremenstova, E., and Luo, L. (1994). Complete transposition requires four active monomers in the Mu transposase tetramer. Genes Dev. 8, 2416-2428,

Berg, D. E. and Howe, M. M. Mobile DNA. (1989). American Society for Microbiology, Washington DC.

Boeke, J. D. and Stoye, J. P. (1997) in Retroviruses, eds. Coffin, J. M., Hughes, S. H. & Varmus, H. E. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 343–435.

Bowman, G.D., O'Donnell, M. and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature 429, 724–730.

Bradley, C.M,. Ronning, D.R., Ghirlando, R., Craigie, R., Dyda, F. (2005). Structural basis for DNA bridging by barrier-to-autointegration factor. Nat. Struct. Mol. Biol. 12: 935-936.

Brenner, S.L., Zlotnick, A., Griffith, J.D. (1988). RecA protein self-assembly. Multiple discrete aggregation states. J. Mol. Biol. 204, 959-972.

Bukhari, A.I. (1977). Transposition of DNA sequences. Brookhaven Symp Biol. 29, 218-232.

Burke, W.D., Malik, H.S., Lathe, W.C. and Eickbush, T.H. (1998). Are retrotransposons long-term hitchhikers? Nature 392, 141-142.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J. and Rance, M. (2010). Protein NMR Spectroscopy: Principles and Practice. Elsevier Science.

Chaconas, G., Gloor, G. and Miller, J.L. (1985). Amplification and purification of the bacteriophage Mu encoded B transposition protein. J. Biol. Chem. 260, 2662–2669.

Chaconas G. (1987) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Chen, Z., Yang, H. and Pavletich, N.P. (2008). Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. Nature 453, 489–484.

Clubb, R,T., Omichinski, J.G., Savilahti ,H., Mizuuchi, K., Gronenborn, A.M., and Clore, G.M. A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase. Structure 15, 1041-1048.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society 117, 5179-5197.

Coros, C.J., Sekino, Y., Baker, T.A. and Chaconas, G. (2003). Effect of mutations in the C-terminal domain of Mu B on DNA binding and interactions with Mu A transposase. J. Biol. Chem. 278, 31210–31217.

Craig, N.L. (1996). V(D)J recombination and transposition: closer than expected. Science 271, 1512-1514.

Craig, N.L. (1997). Target site selection in transposition. Annu. Rev. Biochem. 66, 437–474.

Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (2002) Mobile DNA II. Washington, DC: American Society for Microbiology Press.

Craigie, R., Arndt-Jovin, D.J. and Mizuuchi, K. (1985). A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. Proc Natl Acad Sci USA 82, 7570–7574.

Craigie, R. and Mizuuchi, K. (1985). Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. Cell 41, 867-876.

Craigie, R., Mizuuchi, K., Bushman, F.D., Engelman, A. (1991). A rapid in vitro assay for HIV DNA integration. Nucleic Acid Res. 19, 2729-2734.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. Science 289, 77-85.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of Biomolecular NMR 6, 277-293.

Duderstadt, K.E., Chuang, K. and Berger, J.M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. Nature 478, 209–213.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies DR. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science 266, 1981-1986.

Effantin, G., Ishikawa, T., De Donatis, G.M., Maurizi, M.R. and Steven, A.C. (2010). Local and global mobility in the ClpA AAA+ chaperone detected by cryo-electron microscopy: functional connotations. Structure 18, 553-562.

Egelman, E.H. (2007). The iterative helical real space reconstruction method: Surmounting the problems posed by real polymers. J. Struct. Biol. 157, 83–94.

Erzberger, J.P. and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu. Rev. Biophys. Biomol. Struct. 35, 93–114.

Gay, N.J., Tybulewicz, V.L. and Walker, J.E. (1986). Insertion of transposon Tn7 into the Escherichia coli glmS transcriptional terminator. Biochem J. 234, 111-117.

Ge J, Lou Z, Harshey RM. (2010). Immunity of replicating Mu to self-integration: a novel mechanism employing MuB protein. Mobile DNA 1, 1-8.

Goff, S. A. et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296, 92–100.

Goto, N., Mochizuki, A., Inagaki, Y., Horiuchi, S., Tanaka, T., Nakaya, R. (1987). Identification of the DNA sequence required for transposition immunity of the gamma delta sequence. J Bacteriol. 169, 4388–4390.

Greene, E.C. and Mizuuchi, K. (2004). Visualizing the assembly and disassembly mechanisms of the MuB transposition targeting complex. J. Biol. Chem. 279, 16736–16743.

Greene, E.C. and Mizuuchi, K. (2002a). Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. EMBO J. 21,1477–1486.

Greene, E.C. and Mizuuchi, K. (2002b). Direct observation of single MuB polymers: evidence for a DNA-dependent conformational change for generating an active target complex. Mol. Cell. 9,1079–1089.

Greene, E.C. and Mizuuchi, K. (2002c). Target immunity during Mu DNA transposition. Transpososome assembly and DNA looping enhance MuA-mediated disassembly of the MuB target complex. Mol. Cell. 10, 1367–1378.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R. R. (1988). Clean TOCSY for proton spin system identification in macromolecules. Journal of the American Chemical Society 110, 7870-7872.

Gringauz, E., Orle, K.A., Waddell, C.S. and Craig, N.L. (1988). Recognition of Escherichia coli attTn7 by transposon Tn7: lack of specific sequence requirements at the point of Tn7 insertion. J Bacteriol. 170, 2832-2840.

Grundy, F.J. and Howe, M.M. (1984). Involvement of the invertible G segment in bacteriophage Mu tail fiber biosynthesis. Virology 134, 296–317.

Güntert, P. (2003). Automated NMR protein structure calculation. Progress in Nuclear Magnetic Resonance Spectroscopy 43, 105-125.

Han, Y.W. and Mizuuchi, K. (2010). Phage Mu transposition immunity: protein pattern formation along DNA by a diffusion-ratchet mechanism. Mol. Cell. 39, 48–58.

Hanson, P.I. and Whiteheart, S.W. (2005). AAA+ proteins: Have engine, will work. Nat. Rev. Mol. Cell. Biol. 6, 519-529.

Haren, L., Ton-Hoang, B. & Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. 53, 245–281.

Hauer, B. and Shapiro, J.A. (1984). Control of Tn7 transposition. Molec. Gen. Genet. 194, 149–158.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77, 51-59.

Hui, M.P., Galkin, V.E., Yu, X., Stasiak, A.Z., Stasiak, A., Waldor, M.K. and Egelman, E.H. (2010). ParA2, a Vibrio cholerae chromosome partitioning protein, forms left-handed helical filaments on DNA. Proc. Natl. Acad. Sci. USA 107, 4590–4595.

Hung, L.H., Chaconas, G. and Shaw, G.S. (2000). The solution structure of the C-terminal domain of the Mu B transposition protein. EMBO J. 19, 5625–5634.

Ishikawa, T., Maurizi, M.R. and Steven, A.C. (2004). The N-terminal substrate-binding domain of ClpA unfoldase is highly mobile and extends axially from the distal surface of ClpAP protease. J. Struct. Biol. 146, 180-188.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. The Journal of Chemical Physics 71, 4546-4553.

Kelley, L.A. and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363–371.

Koradi, R., Billeter, M. and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. Journal of Molecular Graphics 14, 51-55.

Koradi, R., Billeter, M. and Güntert, P. (2000). Point-centered domain decomposition for parallel molecular dynamics simulation. Computer Physics Communications 124, 139-147.

Lander, E. S. et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR 8, 477-486.

Lavoie, B.D. and Chaconas, G. (1995). Transposition of phage Mu DNA. Curr. Top. Microbiol. Immunol. 204, 83–102.

Lee, C.H., Bhagwat, A. and Heffron, F. (1983). Identification of a transposon Tn3 sequence required for transposition immunity. Proc. Natl. Acad. Sci. USA 80, 6765–6769.

Lesterlin, C., Ball, G., Schermelleh, L., Sherratt, D.J. (2014). RecA bundles mediate homology pairing between distant sisters during DNA break repair. Nature 506, 249-253.

Leung, P. C., Teplow, D. B., and Harshey, R. M. (1989). Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. Nature 338, 656-658.

Levchenko, I., Yamauchi, M. and Baker T.A. (1997). ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. Genes Dev. 11, 1561-72.

Luginbühl, P., Güntert, P., Billeter, M. and Wüthrich, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. Journal of Biomolecular NMR 8, 136-146.

Manna, D. and Higgins, N.P. (1999). Phage Mu transposition immunity reflects supercoil domain structure of the chromosome. Mol. Microbiol. 32, 595–606.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36, 344-55.

Miller, J.L., Anderson, S.K., Fujita, D.J., Chaconas, G., Baldwin, D.L. and Harshey, R.M. (1984). The nucleotide sequence of the B gene of bacteriophage Mu. Nucleic Acids Res. 12, 8627–8638.

Maxwell, A., Craigie, R. and Mizuuchi, K. (1987). B protein of bacteriophage mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. Proc. Natl. Acad. Sci. USA 84, 699–703.

May EW, Craig NL. (1996). Switching from cut-and-paste to replicative Tn7 transposition. Science 272, 401–404.

Mendiola, M. V., Bernales, I. and de la Cruz, F. (1994). Differential roles of the transposon termini in IS91 transposition. Proc. Natl Acad. Sci. USA 91, 1922–1926.

Milot, E., Belmaaza, A., Rassart, E., and Chartrand P. (1994). Association of a Host DNA Structure with Retrovital Integration Sites in Chromosomal DNA. Virology 201, 408–412.

Mizuuchi, K. (1983). In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. Cell 35, 785-794.

Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. 61, 1011–1051.

Mizuuchi, M., Baker, T. A., Mizuuchi, K. (1995). Assembly of phage Mu transpososomes: cooperative transitions assisted by protein and DNA scaffolds. Cell 83, 375-385.

Montaño S.P., Pigli, Y.Z. and Rice, P.A. (2012). The Mu transpososome structure sheds light on DDE recombinase evolution. Nature 491, 413–417.

Moran, J. V., DeBerardinis, R. J. and Kazazian, H. H. Jr. (1999). Exon shuffling by L1 retrotransposition. Science 283, 1530–1534.

Morgan GJ, Hatfull GF, Casjens S, Hendrix RW. (2002). Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. J. Mol. Biol. 317, 337-359.

Mumenthaler, C., Güntert, P., Braun, W. and Wüthrich, K. (1997). Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. Journal of Biomolecular NMR 10, 351-362.

Nakayama, C., Teplow, D. B., and Hershey, R. M. (1987). Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. Proc. Natl. Acad. Sci. USA 64, 1809-1813.

Pribil, P.A and Haniford, D.B. (2003) Target DNA bending is an important specificity determinant in target site selection in Tn10 transposition. J Mol Biol 330, 247-259.

Sarnovsky RJ, May EW, Craig NL. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. 15, 6348-6361.

Penefsky, H.S. (1977). Reversible binding of Pi by beef heart mitochondrial adenosine triphosphatase. J. Biol. Chem. 252, 2891–2899.

Piotto, M., Saudek, V. and Sklenár, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. 2, 661-665.

Ramón-Maiques, S., Britton, H.G. and Rubio, V. (2002). Molecular physiology of phosphoryl group transfer from carbamoyl phosphate by a hyperthermophilic enzyme at low temperature. Biochemistry 41, 3916–3924.

Rappas, M., Schumacher, J., Beuron, F., Niwa, H., Bordes, P., Wigneshweraraj, S., Keetch, C.A., Robinson, C.V., Buck, M. and Zhang, X. (2005). Structural insights into the activity of enhancer-binding proteins. Science 307, 1972–1975.

Rappas, M., Schumacher, J., Niwa, H., Buck, M. and Zhang, X. (2006). Structural basis of the nucleotide driven conformational changes in the AAA+ domain of transcription activator PspF. J. Mol. Biol. 357, 481–492.

Rice, P. and Mizuuchi, K. (1995). Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. Cell 82, 209-220.

Robinson, M.K., Bennett, P.M. and Richmond, M.H. (1977) Inhibition of TnA translocation by TnA. J. Bacteriol. 129, 407–14.

Saha, R.P., Lou, Z., Meng, L. and Harshey RM. (2013). Transposable prophage Mu is organized as a stable chromosomal domain of E. coli. PLoS Genet. 9, e 1003902.

Sandmeyer, S.B., Bilanchone, V.W., Clark, D.J., Morcos, P., Carle, G.F. and Brodeur G.M. (1988). Sigma elements are position-specific for many different yeast tRNA genes. Nucleic Acids Res. 16, 1499-14515.

Schumaker S, Clubb RT, Cai M, Mizuuchi K, Gronenborn AM and Clore GM (1997). Solution structure of the Mu end DNA-binding Iβ subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. EMBO J, 16, 7532–7541.

Scott, K.K. (2002). The role of conjugative transposons in spreading antibiotic resistance between bacteria that inhabit the gastrointestinal tract. Cell Mol Life Sci. 59, 2071-2082.

Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.4.1.

Shapiro, J. 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA 76, 1933-1937.

Shapiro, J.A. (2010). Mobile DNA and evolution in the 21st century. Mobile DNA 1, 4–18.

Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. Journal of Biomolecular NMR 44, 213-223.

Söding, J., Biegert, A. and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33, W244–W248.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. protein Exp. Purif. 41, 207-234.

Surette, M. G., Harkness, T., and Chaconas, G. (1991). Stimulation of the Mu A protein-mediated strand cleavage reaction by the Mu B protein, and the requirement of DNA nicking for stable type 1 transpososome formation: in vitro transposition characteristics of mini-Mu plasmids carrying terminal base pair mutations. J. Biol. Chem. 266, 3118-3124.

Suzuki, Y. and Craigie, R. (2002). Regulatory mechanisms by which barrier-to-autointegration factor blocks autointegration and stimulates intermolecular integration of Moloney murine leukemia virus preintegration complexes. J. Virol. 76, 12376-12380.

Teplow, D.B., Nakayama, C., Leung, P.D. and Harshey, R.M. (1988). Structure-function relationships in the transposition protein B of bacteriophage Mu. J. Biol. Chem. 263, 10851–10857.

Van Gent DC, Mizuuchi K, Gellert M. 1996. Similarities between initiation of V(D)J recombination and retroviral integration. Science 271, 1592-1594.

Vecchiarelli, A.G., Han, Y.W., Tan, X., Mizuuchi, M., Ghirlando, R., Biertümpfel, C., Funnell, B.E. and Mizuuchi, K. (2010). ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition. Mol. Microbiol. 78, 78–91.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59, 687-696.

Vuister, G. W. and Bax, A. (1994). Measurement of four-bond HN-H alpha J-couplings in Staphylococcal nuclease. Journal of Biomolecular NMR 4, 193-200.

Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

Waterston, R. H. et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Watson J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M. and Losick, R. (2004) Molecular Biology of the Gene. CSHL Press. 310-336.

Wendler, P., Ciniawsky, S., Kock, M. and Kube, S. (2012). Structure and function of the AAA+ nucleotide binding pocket. BBA 1823, 2–14.

Wu, Z., and Chaconas, G. (1994). Characterization of a region in phage Mu transposase that is involved in interaction with the Mu B protein. J. Biol. Chem. 269, 28829-28833.

Yamauchi, M. and Baker, T.A. (1998). An ATP-ADP switch in MuB controls progression of the Mu transposition pathway. EMBO J. 17, 5509–5518.

Yanagihara, K. and Mizuuchi, K. (2002). Mismatch-targeted transposition of Mu: a new strategy to map genetic polymorphism. Proc. Natl. Acad. Sci. USA 99, 11317–11321.

Zheng R, Ghirlando R, Lee MS, Mizuuchi K, Krause M, Craigie R (2000). Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. Proc. Natl. Acad. Sci. USA 97, 8997-9002.